Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/006603

International filing date: 28 February 2005 (28.02.2005)

Document type: Certified copy of priority document

Document details:
Country/Office: US
Number: 60/547,969
Filing date: 26 February 2004 (26.02.2004)

Date of receipt at the International Bureau: 18 April 2005 (18.04.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1303220

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

March 31, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/547,969*
FILING DATE: *February 26, 2004*
RELATED PCT APPLICATION NUMBER: *PCT/US05/06603*

UNITED STATES PATENT AND TRADEMARK OFFICE
DEPARTMENT OF COMMERCE

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

PTO/SB/16 (08-03)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

Express Mail Label No. EV 327707285 US

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Bernhardt L.	Trout	Cambridge, MA 02139
Daniel I.C.	Wang	Newton, MA 02459

☒ *Additional inventors are being named on the* 1 *separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

Solution Additives for the Attenuation of Protein Aggregation

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☒ Customer Number 25181

OR

□ Firm *or* Individual Name					
Address					
Address					
City		State		ZIP	
Country		Telephone		Fax	

ENCLOSED APPLICATION PARTS ***(check all that apply)***

☒ Specification *Number of Pages* 53

☒ Drawing(s) *Number of Sheets* 13

□ Application Data Sheet. See 37 CFR 1.76

□ CD(s), Number ____

☒ Other (specify) A check in the amount of $160.00 and a return-receipt postcard.

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

□ Applicant claims small entity status. See 37 CFR 1.27.

☒ A check or money order is enclosed to cover the filing fees

☒ The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 06-1448

□ Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($) 160.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

□ Yes, the name of the U.S. Government agency and the Government contract number are: _____.

Respectfully submitted,
SIGNATURE Dana Gordon

Date 2/26/04

TYPED or PRINTED NAME Dana M. Gordon

REGISTRATION NO. 44,719
(if appropriate)

Docket Number: **MTV-073.60**

TELEPHONE 617-832-1000

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SENO FEES OR COMPLETEO FORMS TO THIS AOORESS. SEND TO: Mail Stop Provisional Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PROVISIONAL APPLICATION COVER SHEET

Additional Page

PTO/SB/16 (08-03)
Approved for use through 07/31/2006. OMB 0651-0032
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Docket Number MTV-073.60

INVENTOR(S)/APPLICANT(S)

Given Name (first and middle [if any])	Family or Surname	Residence (City and either State or Foreign Country)
Brian N.	Baynes	Cambridge, MA 02139

[Page 2 of 2]

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

Certificate of Express Mail

I, Shirine Darvish, do hereby certify that the foregoing documents are being deposited with the United States Postal Service as Express Mail, postage prepaid, "Post Office to Addressee", in an envelope addressed to Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria VA 22313-1450 on this date of February 26, 2004.

Shirine Darvish
Express Mail Label: EV 327707285 US
Date of Deposit: February 26, 2004

Submission consisting of:

1. Provisional Application for Patent Cover Sheet (2 pages);
2. Certificate of Express Mailing (1 page);
3. Specification (53 pages);
4. Sheets of Drawings (13 pages);
5. A check in the amount of $160.00 to cover the filing fee; and
6. This return-receipt postcard.

Attorney Docket No.: MTV-073.60

PROVISIONAL APPLICATION FOR PATENT

Solution Additives for the Attenuation of Protein Aggregation

Background of the Invention

The process of protein folding is complex, and a complete understanding of it is one of the grand challenges facing contemporary biochemists. This complexity arises in part from the fact that a nascent protein may not simply fold into its native state under the influence of only the primary solvent (water), but may also interact with other molecules in solution. The effects of other molecules may be favorable for folding, as is the case for molecules like folding chaperones, or unfavorable, as is the case for other partially-unfolded protein molecules.

One of the primary driving forces in protein folding is the burial of exposed hydrophobic residues. Dill, K. *A. Biochemistry* **1990**, 29, 7133-7155. Aggregation results if the hydrophobic collapse should occur in an intermolecular instead of an intramolecular fashion. Because aggregation occurs as a parallel reaction to proper folding, there is kinetic competition between the two pathways. Orsini, G.; Goldberg, VI. E. J. *Biol. Chem.* **1978**, 253, 3453-3458; Zettlmeissl, G.; Rudolph; R.; Jaenicke. R. *Biochemistry* **1979**, 18, 5567-5571; Kiefllaber, T.; Rudolph; R.; Kohler, H.-H.; Buchner, J. *Bio/Technology* **1991**, 9, 825-829; Hevehan, D. L.; Clark, E. D. B. *Biotechnol. Bioeng.* **1997**, 54, 221-230.

Aggregation of misfolded proteins is a significant problem both *in vivo* and *in vitro.* Aggregation has been implicated in human diseases, such as Huntington's, Alzheimer's, and Parkinson's Diseases. Taylor, J. P.; Hardy, J.; Fischbeck; K. H. *Science* **2002**, 296, 1991-1995. In applied biotechnology, aggregation is a significant side reaction of protein refolding, which is an important step in the production of many recombinant proteins. De Bernandez Clark, E.; Schwarz, E.; Rudolph, R. *Methods Enzymol.* **1999**, 309, 217-236.

Both nature and man have developed strategies to combat aggregation. Chaperonins, such as the GroEL/GroES system, surround and isolate partially-folded proteins in the bulk cytosol so they can continue to fold without aggregating. Hartl, F. U.; Hayer-Hartl, M. *Science*

2003, 295, 1852-1858. Similarly, additives to deter aggregation are often included in protein refolding buffers and other *in vitro* applications, such as pharmaceutical formulations. Wang, W. *Int.* J. *Pharm.* **1999**, 185, 129-188.

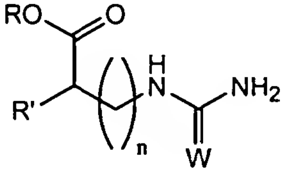
Summary of the Invention

Presently disclosed are classes of additives that, when added to protein solutions, attenuate the rate of aggregation. The members of the classes have two key, well-defined properties that result in their ability to slow aggregation. The present invention also recognizes that there are many other such molecules that have never been synthesized and have never been used to stabilize proteins. Some of them exemplify the two properties above to a greater extent and are, therefore, superior solution additives for the attenuation of protein aggregation.

In one embodiment the present invention relates to a compound comprising a non-protein-binding moiety (NPBM) and at least one protein binding group (PBG). In a further embodiment, the NPBM is a polyol, sugar, amino acid, or dendrimer moiety. In a further embodiment, the polyol moiety is a sorbitol or mannitol moiety. In a further embodiment, the sugar moiety is a glucose, sucrose, or trehalose moiety. In a further embodiment, the amino acid moiety is an arginine betaine, proline, or ectoine moiety. In a further embodiment, the dendrimer moiety is based on benzene, pentaerythritol, P(CH_2OH)$_3$, or TRIS.

In a further embodiment, the PBG is a urea, guanidinium ion, detergent, amino acid, denaturant, surfactant, polysorbate, polaxamer, citrate, chaotrope, or acetate group. In a further embodiment, the PBG is a guanidinium ion. In a further embodiment, the PBG is sodium dodecyl sulfate.

In another embodiment, the present invention relates to a compound of formula **I**:



I

wherein:

R is an electron pair, H, alkyl, aryl, heteroaryl, aralkyl, heteroaralkyl, or an alkali metal;

R' is H, alkyl, aryl, heteroaryl, aralkyl, heteroaralkyl, or R"H_2N;

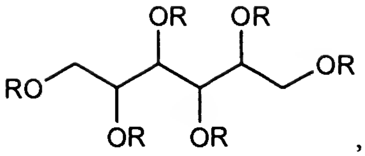
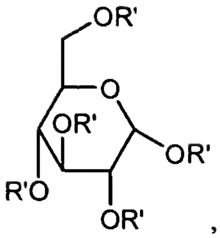
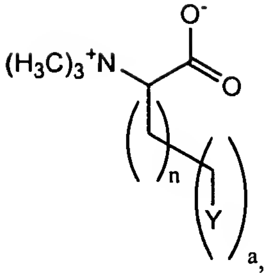
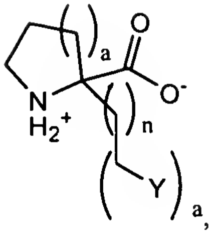
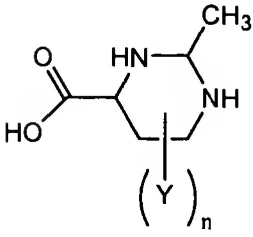
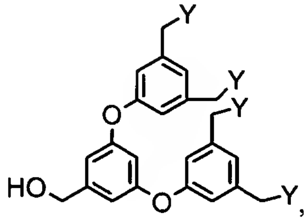
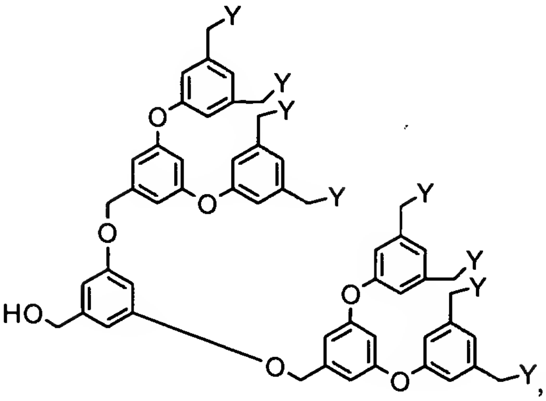
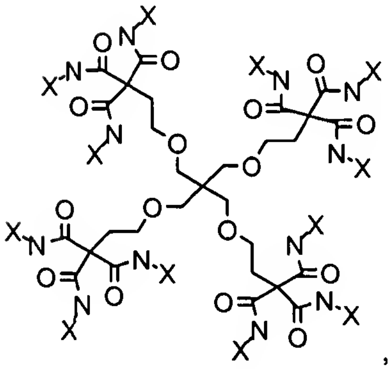
R" is an electron pair, H, alkyl, aryl, heteroaryl, aralkyl, or heteroaralkyl;

W is O, NH_2^+(halogen)$^-$, or S; and

n is 1, 2, or 4-100.

In a further embodiment, the present invention relates to a compound of formula **I** and the attendant definitions, wherein R is an electron pair. In a further embodiment, R' is H. In a further embodiment, R' is R"H_2N. In a further embodiment, R' is H_3N^+. In a further embodiment, W is $NH_2^+Cl^-$. In a further embodiment, n is 1. In a further embodiment, n is 2. In a further embodiment, n is 4. In a further embodiment, n is 5. In a further embodiment, n is 6. In a further embodiment, R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 1. In a further embodiment, R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 2. In a further embodiment, R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 4. In a further embodiment, R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 5. In a further embodiment, R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 6. In a further embodiment, R is an electron pair, R' is H_3N^+, W is O, and n is 1. In a further embodiment, R is an electron pair, R' is H_3N^+, W is O, and n is 2. In a further embodiment, R' is H_3N^+, W is O, and n is 4. In a further embodiment, R is an electron pair, R' is H_3N^+, W is O, and n is 5. In a further embodiment, R is an electron pair, R' is H_3N^+, W is O, and n is 6. In a further embodiment, R is an electron pair, R' is H, W is $NH_2^+Cl^-$, and n is 1. In a further embodiment, R is an electron pair, R' is H, W is $NH_2^+Cl^-$, and n is 2. In a further embodiment, R is an electron pair, R' is H^+, W is $NH_2^+Cl^-$, and n is 4. In a further embodiment, R is an electron pair, R' is H, W is $NH_2^+Cl^-$, and n is 5. In a further embodiment, R is an electron pair, R' is H, W is $NH_2^+Cl^-$, and n is 6. In a further embodiment, R is an electron pair, R' is H, W is O, and n is 1. In a further embodiment, R is an electron pair, R' is H, W is O, and n is 2. In a further embodiment, R is an electron pair, R' is H, W is O, and n is 4. In a further embodiment, R is an electron pair, R' is H, W is O, and n is 5. In a further embodiment, R is an electron pair, R' is H, W is O, and n is 6.

In another embodiment, the present invention relates to one of the following compounds:


,

,



,


, or

,

wherein, independently for each occurrence,

R is H or CH_2Y;

R' is H, a sugar radical, or CH_2Y;

n is an integer from 1 to 100, inclusive;

a is 1, 2, or 3;

X is $C(CH_2Y)_3$; and

Y is a protein binding group,

wherein at least one Y is present in all compounds.

In a further embodiment, Y is a guanidinium ion.

In another embodiment, the present invention relates to a method of screening compounds for the property of inhibiting protein aggregation in solution, comprising:

a) computing a set of parameters utilizing molecular modeling based on compounds known to have the property of inhibiting protein aggregation;

b) applying those parameters to other compounds; and

c) choosing the compounds that meet the criteria of those parameters.

In another embodiment, the present invention relates to a method of preparing new compounds having the property of protein aggregation inhibition in solution, comprising:

a) computing a set of parameters utilizing molecular modeling based on compounds known to have the property of inhibiting protein aggregation;

b) designing compounds based on those parameters; and

c) synthesizing the compounds.

In another embodiment, the present invention relates to a method of classifying compounds as either inhibitory of protein aggregation in solution or not inhibitory of protein aggregation in solution, comprising:

a) determining the phase space trajectories of the protein, solvent, and additive using molecular dynamics;

b) calculating the distance, r, between the center of mass for both the solvent molecule and additive molecule to the protein's van der Waals surface;

c) determining the minimum distance, r*, at which no significant differences between the local (r = r*) and bulk density are observed;

d) determining which molecules lie within the distance, r*, from the protein surface and classifying these molecules as the local domain;

e) determining which molecules lie outside the distance, r*, from the protein surface and classifying these molecules as the bulk domain;

f) determining the instantaneous preferential binding coefficient, $\Gamma_{XP}(t)$, using the following formula:

$$\Gamma_{XP}(t) = n^{II}_{X} - n^{I}_{X} (n^{II}_{W} / n^{I}_{W})$$

wherein:

n^{II}_{X} = the number of additive molecules in the bulk domain;

n^{I}_{X} = the number of additive molecules in the local domain;

n^{II}_{W} = the number of solvent molecules in the bulk domain; and

n^{I}_{W} = the number of solvent molecules in the local domain; and

g) calculating the preferential binding coefficient, Γ_{XP}, as the time average of each of the values in step f) using the following formula:

$$\Gamma_{XP} = \frac{1}{t}\int_{0}^{t}\Gamma_{XP}(t')dt'.$$

These embodiments of the present invention, other embodiments, and their features and characteristics, will be apparent from the description, drawings and claims that follow.

Brief Description of the Figures

Figure 1 depicts a simplified dimerization reaction-coordinate diagram for the reaction U + U → A_2 (equation 2). The dotted line is the reaction coordinate in water and the solid line is the reaction coordinate in the presence of an additive having the two anti-aggregation properties discussed. Protein molecules are represented by black coils and the additive by dark grey circles. The energy difference between the reactants (U + U) and the transition state determines the rate of the reaction. In the A_2 state, the region between the protein molecules (light grey oval) is preferentially hydrated because water can enter this region but the additive cannot. This preferential hydration increases the free energy of the transition state, increases the energy barrier for the reaction, and slows the reaction rate.

Figure 2 depicts arginine derivatives with shorter (left) and longer (right) methylene linkers between their amino acid backbone and guanidino functional groups.

Figure 3 depicts molecules that will be preferentially-oriented at the protein-solvent interface. Molecule (a) is a derivative of glucose (stabilizer) linked to a dimethyl-guanidino (destabilizer) moiety. Molecule (b) is a polyol (stabilizer) with a guanidino group (destabilizer) attached to one end.

Figure 4 depicts the physical interpretation of the preferential binding coefficient. Interactions of solvent molecules with the protein at the protein-solvent interface generally induce solvent concentration differences in the local (II) and bulk (I) domains. Γ_{XP} is the thermodynamic measure of the number of additive molecules bound to the protein, or in other words, the excess number of additive molecules in the vicinity of the protein versus the number of additive molecules in an equivalent volume of bulk solution.

Figure 5 depicts a simulation cell containing RNase T1 (center spheres) solvated by water (thin lines) and urea (spheres).

Figure 6 depicts radial distribution functions of water, urea, and glycerol shown for simulations of RNase T1 in glycerol and urea solutions (left) and RNase A in a glycerol solution (right). In the left-hand figure, the difference between the two $gw(r)$ functions is not visible at this scale.

Figure 7 depicts apparent preferential binding coefficient as a function of the cutoff distance between the local and bulk domains for simulations of RNase T1 in glycerol and urea solution.

Figure 8 depicts $\Gamma_{xp}(t)$ probability density function. A wide range of values of $\Gamma_{xp}(t)$ are sampled as water and cosolvent molecules diffuse between the local and bulk domains.

Figure 9 depicts the correlation of solvent-accessible area and the number of water molecules in the local domain of constituent groups. Each point represents a constituent group of either a type of amino acid side chain or the protein backbone in one of the three simulations shown in Table 2. The solvent accessible area of a constituent group and the number of water molecules in the local domain of the solvent near the group (n_{wi}) are correlated.

Figure 10 depicts the binding behavior of glycerol and water with the 15 serine residues in RNase T1 as shown in a plot of the number of glycerol molecules in the local domain of each serine residue versus the number of water molecules in the same volume. The labels are the one-letter codes for each amino acid side chain, and "B" is the protein backbone. The line represents the bulk glycerol composition. Ser 17, 35, and 72 have positive preferential binding coefficients, Ser 63 has a negative preferential binding coefficient, and the remaining 11 serine residues have essentially zero values for their preferential binding coefficients.

Figure 11 depicts the local binding behavior of urea and water with the amino acid backbone and side chains in RNase T1. The labels are the one-letter codes for the amino acid side chains, and "B" is the protein backbone. The line denotes the bulk urea concentration. In addition to the protein backbone and Ser, the hydrophobic amino acids Cys, Gly, Leu, Phe, Pro, Tyr, and Val all preferentially bind urea, while the hydrophilic Asp preferentially binds water.

Figure 12 depicts the group preferential binding coefficients for glycerol with the amino acid backbone and side chains in RNase T1. The labels are the one-letter codes for the amino acid side chains, and "B" is the protein backbone. The line denotes the bulk glycerol concentration. Tyr and Gly preferentially bind glycerol; Asp and Glu preferentially bind water; and the binding coefficients of the other groups are not statistically different from zero.

Figure 13 depicts the local binding behavior of glycerol with the amino acid backbone and side chains in RNase A. The labels are the one-letter codes for the amino acid side chains, and "B" is the protein backbone. The line denotes the bulk glycerol concentration. All of the constituent groups in RNase A either preferentially bind water or are neutral.

Detailed Description of the Invention

Definitions

For convenience, before further description of the present invention, certain terms employed in the specification, examples and appended claims are collected here. These definitions should be read in light of the remainder of the disclosure and understood as by a person of skill in the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by a person of ordinary skill in the art.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

The terms "comprise" and "comprising" are used in the inclusive, open sense, meaning that additional elements may be included.

The term “including” is used to mean “including but not limited to”. “Including” and “including but not limited to” are used interchangeably.

The term “additive” as used herein refers to any component other than the subject protein and the main solvent. Non-limiting examples of additives include small molecules, cosolvents, buffer salts, and stabilizers.

The term “dendrimer” is used to mean a broad class of polymers constructed via stepwise polymerization from a central “core unit,” one or more “branching units,” and several “surface units.” The review of Matthews (1998) provides an overview of dendrimers including compositions and synthetic routes. Core units may include (but are not limited to) carbon, nitrogen, phosphorous, benzene, and porphyrins. A non-extensive collection of 17 specific chemistries that are used to create branching units are summarized in Table 2 of Matthews (1998).

The term “TRIS” is art-recognized and refers to tris(hydroxymethyl)aminomethane.

The term “aliphatic” is an art-recognized term and includes linear, branched, and cyclic alkanes, alkenes, or alkynes. In certain embodiments, aliphatic groups in the present invention are linear or branched and have from 1 to about 20 carbon atoms.

The term “alkyl” is art-recognized, and includes saturated aliphatic groups, including straight-chain alkyl groups, branched-chain alkyl groups, cycloalkyl (alicyclic) groups, alkyl substituted cycloalkyl groups, and cycloalkyl substituted alkyl groups. In certain embodiments, a straight chain or branched chain alkyl has about 30 or fewer carbon atoms in its backbone (e.g., C_1-C_{30} for straight chain, C_3-C_{30} for branched chain), and alternatively, about 20 or fewer. Likewise, cycloalkyls have from about 3 to about 10 carbon atoms in their ring structure, and alternatively about 5, 6 or 7 carbons in the ring structure.

Unless the number of carbons is otherwise specified, “lower alkyl” refers to an alkyl group, as defined above, but having from one to ten carbons, alternatively from one to about six carbon atoms in its backbone structure. Likewise, “lower alkenyl” and “lower alkynyl” have similar chain lengths.

The term “aralkyl” is art-recognized, and includes alkyl groups substituted with an aryl group (e.g., an aromatic or heteroaromatic group).

The terms “alkenyl” and “alkynyl” are art-recognized, and include unsaturated aliphatic groups analogous in length and possible substitution to the alkyls described above, but that contain at least one double or triple bond respectively.

The term “heteroatom” is art-recognized, and includes an atom of any element other than carbon or hydrogen. Illustrative heteroatoms include boron, nitrogen, oxygen, phosphorus, sulfur and selenium, and alternatively oxygen, nitrogen or sulfur.

The term “aryl” is art-recognized, and includes 5-, 6- and 7-membered single-ring aromatic groups that may include from zero to four heteroatoms, for example, benzene, naphthalene, anthracene, pyrene, pyrrole, furan, thiophene, imidazole, oxazole, thiazole, triazole, pyrazole, pyridine, pyrazine, pyridazine and pyrimidine, and the like. Those aryl groups having heteroatoms in the ring structure may also be referred to as “heteroaryl” or “heteroaromatics.” The aromatic ring may be substituted at one or more ring positions with such substituents as described above, for example, halogen, azide, alkyl, aralkyl, alkenyl, alkynyl, cycloalkyl, hydroxyl, alkoxyl, amino, nitro, sulfhydryl, imino, amido, phosphonate, phosphinate, carbonyl, carboxyl, silyl, ether, alkylthio, sulfonyl, sulfonamido, ketone, aldehyde, ester, heterocyclyl, aromatic or heteroaromatic moieties, -CF_3, -CN, or the like. The term “aryl” also includes polycyclic ring systems having two or more cyclic rings in which two or more carbons are common to two adjoining rings (the rings are “fused rings”) wherein at least one of the rings is aromatic, e.g., the other cyclic rings may be cycloalkyls, cycloalkenyls, cycloalkynyls, aryls and/or heterocyclyls.

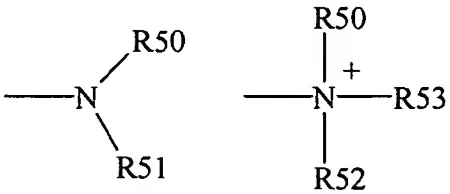
The terms ortho, meta and para are art-recognized and apply to 1,2-, 1,3- and 1,4-disubstituted benzenes, respectively. For example, the names 1,2-dimethylbenzene and ortho-dimethylbenzene are synonymous.

The terms “heterocyclyl” and “heterocyclic group” are art-recognized, and include 3- to about 10-membered ring structures, such as 3- to about 7-membered rings, whose ring structures include one to four heteroatoms. Heterocycles may also be polycycles. Heterocyclyl groups include, for example, thiophene, thianthrene, furan, pyran, isobenzofuran, chromene, xanthene, phenoxathiin, pyrrole, imidazole, pyrazole, isothiazole, isoxazole, pyridine, pyrazine, pyrimidine, pyridazine, indolizine, isoindole, indole, indazole, purine, quinolizine, isoquinoline, quinoline, phthalazine, naphthyridine, quinoxaline, quinazoline, cinnoline, pteridine, carbazole, carboline, phenanthridine, acridine, pyrimidine, phenanthroline, phenazine, phenarsazine, phenothiazine, furazan, phenoxazine, pyrrolidine, oxolane, thiolane, oxazole, piperidine, piperazine, morpholine, lactones, lactams such as azetidinones and pyrrolidinones, sultams, sultones, and the like. The heterocyclic ring may be substituted at one or more positions with such substituents as described above, as for example, halogen, alkyl, aralkyl, alkenyl, alkynyl, cycloalkyl, hydroxyl, amino, nitro, sulfhydryl, imino, amido, phosphonate, phosphinate, carbonyl, carboxyl, silyl, ether, alkylthio, sulfonyl, ketone, aldehyde, ester, a heterocyclyl, an aromatic or heteroaromatic moiety, $-CF_3$, -CN, or the like.

The terms "polycyclyl" and "polycyclic group" are art-recognized, and include structures with two or more rings (e.g., cycloalkyls, cycloalkenyls, cycloalkynyls, aryls and/or heterocyclyls) in which two or more carbons are common to two adjoining rings, e.g., the rings are "fused rings". Rings that are joined through non-adjacent atoms, e.g., three or more atoms are common to both rings, are termed "bridged" rings. Each of the rings of the polycycle may be substituted with such substituents as described a bove, as for example, halogen, alkyl, aralkyl, alkenyl, alkynyl, cycloalkyl, hydroxyl, amino, nitro, sulfhydryl, imino, amido, phosphonate, phosphinate, carbonyl, carboxyl, silyl, ether, alkylthio, sulfonyl, ketone, aldehyde, ester, a heterocyclyl, an aromatic or heteroaromatic moiety, $-CF_3$, -CN, or the like.

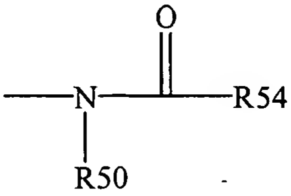
The term "carbocycle" is art-recognized and includes an aromatic or non-aromatic ring in which each atom of the ring is carbon. The flowing art-recognized terms have the following meanings: "nitro" means $-NO_2$; the term "halogen" designates -F, -Cl, -Br or -I; the term "sulfhydryl" means -SH; the term "hydroxyl" means -OH; and the term "sulfonyl" means $-SO_2^-$.

The terms "amine" and "amino" are art-recognized and include both unsubstituted and substituted amines, e.g., a moiety that may be represented by the general formulas:



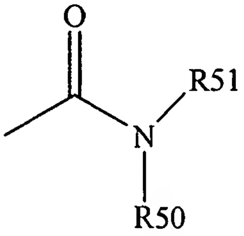
wherein R50, R51 and R52 each independently represent a hydrogen, an alkyl, an alkenyl, $-(CH_2)_m$-R61, or R50 and R51, taken together with the N atom to which they are attached complete a heterocycle having from 4 to 8 atoms in the ring structure; R61 represents an aryl, a cycloalkyl, a cycloalkenyl, a heterocycle or a polycycle; and m is zero or an integer in the range of 1 to 8. In certain embodiments, only one of R50 or R51 may be a carbonyl, e.g., R50, R51 and the nitrogen together do not form an imide. In other embodiments, R50 and R51 (and optionally R52) each independently represent a hydrogen, an alkyl, an alkenyl, or -$(CH_2)_m$-R61. Thus, the term "alkylamine" includes an amine group, as defined above, having a substituted or unsubstituted alkyl attached thereto, i.e., at least one of R50 and R51 is an alkyl group.

The term "acylamino" is art-recognized and includes a moiety that may be represented by the general formula:



wherein R50 is as defined above, and R54 represents a hydrogen, an alkyl, an alkenyl or -$(CH_2)_m$-R61, where m and R61 are as defined above.

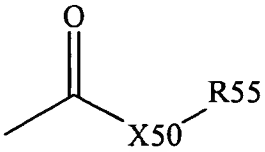
The term "amido" is art-recognized as an amino-substituted carbonyl and includes a moiety that may be represented by the general formula:

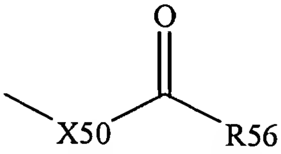


wherein R50 and R51 are as defined above. Certain embodiments of the amide in the present invention will not include imides which may be unstable.

The term "alkylthio" is art-recognized and includes an alkyl group, as defined above, having a sulfur radical attached thereto. In certain embodiments, the "alkylthio" moiety is represented by one of -S-alkyl, -S-alkenyl, -S-alkynyl, and -S-$(CH_2)_m$-R61, wherein m and R61 are defined above. Representative alkylthio groups include methylthio, ethyl thio, and the like.

The term "carbonyl" is art-recognized and includes such moieties as may be represented by the general formulas:

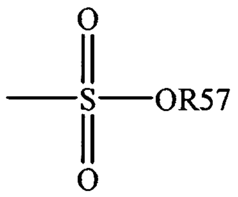




wherein X50 is a bond or represents an oxygen or a sulfur, and R55 represents a hydrogen, an alkyl, an alkenyl, -$(CH_2)_m$-R61or a pharmaceutically acceptable salt, R56 represents a hydrogen, an alkyl, an alkenyl or $-(CH_2)_m$-R61, where m and R61 are defined above. Where X50 is an oxygen and R55 or R56 is not hydrogen, the formula represents an "ester". Where X50 is an oxygen, and R55 is as defined above, the moiety is referred to herein as a carboxyl group, and particularly when R55 is a hydrogen, the formula represents a "carboxylic acid". Where X50 is an oxygen, and R56 is hydrogen, the formula represents a "formate". In general, where the oxygen atom of the above formula is replaced by sulfur, the formula represents a "thiocarbonyl" group. Where X50 is a sulfur and R55 or R56 is not hydrogen, the formula represents a "thioester." Where X50 is a sulfur and R55 is hydrogen, the formula represents a "thiocarboxylic acid." Where X50 is a sulfur and R56 is hydrogen, the formula represents a "thioformate." On the other hand, where X50 is a bond, and R55 is not hydrogen, the above formula represents a "ketone" group. Where X50 is a bond, and R55 is hydrogen, the above formula represents an "aldehyde" group.

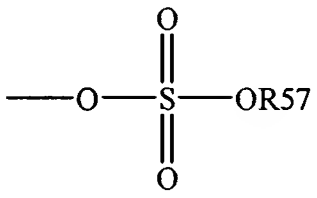
The terms "alkoxyl" or "alkoxy" are art-recognized and include an alkyl group, as defined above, having an oxygen radical attached thereto. Representative alkoxyl groups include methoxy, ethoxy, propyloxy, tert-butoxy and the like. An "ether" is two hydrocarbons covalently linked by an oxygen. Accordingly, the substituent of an alkyl that renders that alkyl an ether is or resembles an alkoxyl, such as may be represented by one of -O-alkyl, -O-alkenyl, -O-alkynyl, -O-$(CH_2)_m$-R61, where m and R61 are described above.

The term "sulfonate" is art-recognized and includes a moiety that may be represented by the general formula:



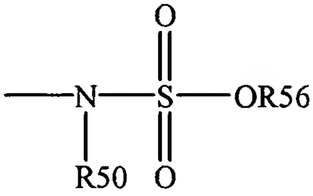
in which R57 is an electron pair, hydrogen, alkyl, cycloalkyl, or aryl.

The term "sulfate" is art-recognized and includes a moiety that may be represented by the general formula:



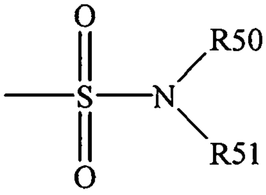
in which R57 is as defined above.

The term “sulfonamido” is art-recognized and includes a moiety that may be represented by the general formula:



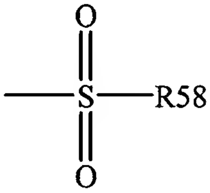
in which R50 and R56 are as defined above.

The term “sulfamoyl” is art-recognized and includes a moiety that may be represented by the general formula:



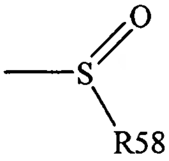
in which R50 and R51 are as defined above.

The term “sulfonyl” is art-recognized and includes a moiety that may be represented by the general formula:



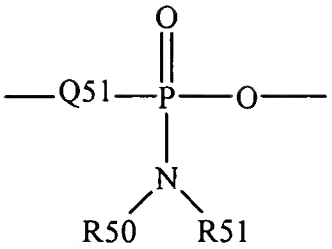
in which R58 is one of the following: hydrogen, alkyl, alkenyl, alkynyl, cycloalkyl, heterocyclyl, aryl or heteroaryl.

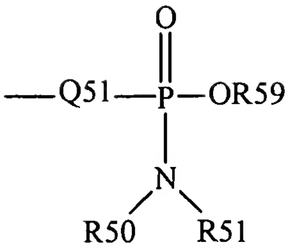
The term “sulfoxido” is art-recognized and includes a moiety that may be represented by the general formula:



in which R58 is defined above.

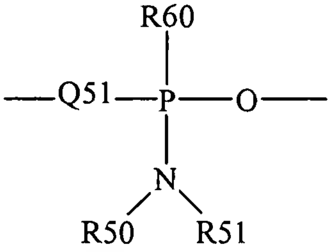
The term “phosphoramidite” is art-recognized and includes moieties represented by the general formulas:

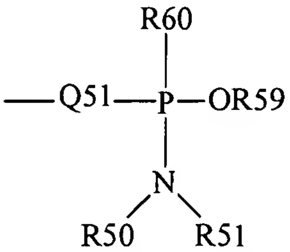




wherein Q51, R50, R51 and R59 are as defined above.

The term "phosphonamidite" is art-recognized and includes moieties represented by the general formulas:





wherein Q51, R50, R51 and R59 are as defined above, and R60 represents a lower alkyl or an aryl.

Analogous substitutions may be made to alkenyl and alkynyl groups to produce, for example, aminoalkenyls, aminoalkynyls, amidoalkenyls, amidoalkynyls, iminoalkenyls, iminoalkynyls, thioalkenyls, thioalkynyls, carbonyl-substituted alkenyls or alkynyls.

The definition of each expression, e.g. alkyl, m, n, etc., when it occurs more than once in any structure, is intended to be independent of its definition elsewhere in the same structure unless otherwise indicated expressly or by the context.

For purposes of this invention, the chemical elements are identified in accordance with the Periodic Table of the Elements, CAS version, Handbook of Chemistry and Physics, 67th Ed., 1986-87, inside cover.

Overview

Proteins are widely used in medical and industrial applications. One of the major difficulties encountered in these applications is that proteins are prone to degradation by a variety of routes, the most common of which is aggregation. Aggregated protein generally does not have the same functionality as normal, native protein. For example, in pharmaceutical applications, the consequences of administering aggregated drug to a patient can be severe because aggregates can be cytotoxic; and they generally induce an immune response. Bucciatini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigh, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. *Nature* **2002,** *416,* 507-511; Braun, A.; Kwee, L.; Labow, M. A.; Alsenz, J. Pharm. Res. **1997**, 14, 1472-1478. Due to these and other negative effects, protein solutions often contain one or more additives designed to deter aggregation. Wang, W. Int. J. Pharm. **1999**, 185,129-188.

As a protein folds, there is kinetic competition between the proper folding pathway and side reactions, such as aggregation. This is depicted schematically in equations 1 and 2:

$$U \rightarrow N \qquad (1)$$

$$U + U \rightarrow A_2 \qquad (2)$$

where U represents an unfolded protein; N represents a folded, native protein; and A_2 represents a small aggregate species.

Alternatively, if the protein is initially in its native state, such as in a pharmaceutical formulation, aggregation proceeds through formation of a partially-unfolded intermediate, I, which can aggregate in a sense analogous to an unfolded protein:

$$N \rightleftharpoons I \qquad (3)$$

$$I + I \rightarrow A_2 \qquad (4)$$

For industrial and medical applications, it is desirable to eliminate or minimize the formation of protein aggregates. In protein folding or refolding processes, decreasing the rate of aggregation results in a higher yield of active, properly-folded protein. In pharmaceutical formulations, decreasing the rate of aggregation causes more drug to remain in its active form and eliminates the possibly dangerous side effects of administering aggregated protein to the patient. To minimize aggregation, various conditions, such as temperature, pH, and the type and amount of buffer additives, are screened experimentally to identify an optimum set of conditions.

Additives that attenuate protein aggregation have historically been identified by heuristic, experimental screening methods. Identification of specific additive-protein combinations that result in decreased aggregation via these techniques has led to little fundamental understanding about what properties of a molecule lead to the ability to suppress protein aggregation. This gap in understanding has prevented development of rational strategies to prevent protein aggregation.

Through the mechanistic understanding summarized presently, two fundamental properties of a good anti-aggregation additive have been identified. This discovery allows additives to be selected based on their relative ranking in terms of these two properties, thus narrowing experimental testing to molecules likely to have optimal performance. It also enables molecules to be classified based on whether they may have the ability to attenuate aggregation. The rational, mechanistic classification schemes of the present invention will allow entire classes of protein-aggregation-attenuating additives and formulations to be identified.

Additionally, a quantitative method based on molecular dynamics simulations using all atom potential models has been developed and validated for calculating preferential binding coefficients. The present invention is not a derivative of thermodynamic integration or thermodynamic perturbation methods and requires only a single trajectory to compute the transfer free energy of a protein into a weak-binding additive system. The results match experimental data well for glycerol and urea solutions, covering a range of positive and negative binding behavior. The present invention also augments experimentally-observable, macroscopic thermodynamics with the mechanistic insight provided by a molecular-level, statistical mechanical model.

Variations in the radial distribution functions with distance for each additive are evident up to about 6Å, i.e., roughly two solvation shells of water, away from the protein. Glycerol is not totally excluded from close contact with the protein, but glycerol is less likely than urea to be found in such a position. The radial distribution functions of water and additives are sufficient to calculate preferential binding coefficients by integrating over a suitable solvent volume.

The binding behavior of the amino acid side chains in RNase T1 qualitatively follow a hydrophilic series, with more hydrophilic amino acids in the protein tending to have a higher concentration of water in their vicinity. The constituent group binding behavior differs between the groups in RNase A to those in RNase Tl. Development of a group contribution method at the amino acid level for estimating binding coefficients or transfer free energies of whole proteins is complicated by the wide range of coordination behaviors observed for single types of amino acids in different environments on the protein surface.

In the pharmaceutical industry, many protein drugs are synthesized in bacterial hosts, such as E. *coli*, in the form of solid, partially-aggregated precipitates called inclusion bodies. These inclusion bodies must be unfolded and solubilized, and then refolded to form active protein. During refolding, proteins are especially susceptible to aggregation, and additives must be used to minimize aggregation and increase the yield of biologically-active protein. The compounds of the present invention are ideal for use in these circumstances because they will slow the rate of aggregation and therefore increase the yield of active protein. Likewise, when pharmaceutically-active proteins are formulated in aqueous solution, additives are used to prevent aggregation during storage, thereby increasing its shelf-life. The compounds of the present invention are also useful in preventing aggregation in these circumstances. Additional applications can be envisioned by those of ordinary skill in the art of protein stabilization. The above applications are meant to be only exemplary and not limiting in any way.

Select Preferred Embodiments

In a preferred embodiment, the present invention relates to a method of suppressing or preventing aggregation of a protein in solution, comprising the step of combining in a solution a compound of the present invention and a protein. In certain embodiments, the protein is a recombinant protein. In certain embodiments, the protein is a recombinant antibody. In certain embodiments, the protein is a recombinant human antibody. In certain embodiments, the protein is a recombinant mammalian protein. In certain embodiments, the protein is a recombinant human protein. In certain embodiments, the protein is recombinant human insulin, recombinant human erythropoietin or a recombinant human interferon. In certain embodiments, the solution is an aqueous solution. In certain embodiments, the protein is a recombinant protein; and the solution is an aqueous solution. In certain embodiments, the protein is a recombinant human antibody; and the solution is an aqueous solution. In certain embodiments, the protein is a recombinant human protein; and the solution is an aqueous solution.

In a preferred embodiment, the present invention relates to a method of suppressing or preventing aggregation of a protein in solution, comprising the step of combining in a solution a compound of the present invention and a protein. In certain embodiments, the protein is a recombinant protein. In certain embodiments, the protein is a recombinant antibody. In certain embodiments, the protein is a recombinant human antibody. In certain embodiments, the protein is a recombinant mammalian protein. In certain embodiments, the protein is a recombinant human protein. In certain embodiments, the protein is recombinant human insulin, recombinant human erythropoietin or a recombinant human interferon. In certain embodiments, the solution is an aqueous solution. In certain embodiments, the protein is a recombinant protein; and the solution is an aqueous solution. In certain embodiments, the protein is a recombinant human antibody; and the solution is an aqueous solution. In certain embodiments, the protein is a recombinant human protein; and the solution is an aqueous solution.

In a third preferred embodiment, the present invention relates to a method of decreasing the toxicological risk associated with administering a protein to a mammal in need thereof, comprising the steps of adding to a first solution of a protein a compound of the present invention to give a second solution; and administering to a mammal in need thereof a therapeutic amount of said second solution. In certain embodiments, the protein is a recombinant protein. In certain embodiments, the protein is a recombinant antibody. In certain embodiments, the protein is a recombinant human antibody. In certain embodiments, the protein is a recombinant mammalian protein. In certain embodiments, the protein is a recombinant human protein. In certain embodiments, the protein is recombinant human insulin, recombinant human erythropoietin or a recombinant human interferon. In certain embodiments, the first solution and the second solution are aqueous solutions. In certain embodiments, the protein is a recombinant protein; and the first solution and the second solution are aqueous solutions. In certain embodiments, the protein is a recombinant human antibody; and the first solution and the second solution are aqueous solutions. In certain embodiments, the protein is a recombinant human protein; and the first solution and the second solution are aqueous solutions.

In another preferred embodiment, the present invention relates to a method of facilitating native folding of a recombinant protein in solution, comprising the step of combining in a solution a compound of the present invention and a recombinant protein. In certain embodiments, the recombinant protein is a recombinant antibody. In certain embodiments, the recombinant protein is a recombinant human antibody. In certain embodiments, the recombinant protein is a recombinant mammalian protein. In certain embodiments, the recombinant protein is a recombinant human protein. In certain embodiments, the recombinant protein is recombinant human insulin, recombinant human erythropoietin or a recombinant human interferon. In certain embodiments, the solution is an aqueous solution. In certain embodiments, the recombinant protein is a recombinant human antibody; and the solution is an aqueous solution. In certain embodiments, the recombinant protein is a recombinant human protein; and the solution is an aqueous solution.

Kinetic model approach for stabilizing proteins towards aggregation

To see how additives affect aggregation rate, the rate constant for aggregation, k_{agg}, can be expressed using transition state theory as:

$$k_{agg} = \frac{k_b T}{h} K^{\ddagger} \qquad (5)$$

where k_b is Boltzmann's constant, T is the absolute temperature, h, is Planck's constant, and $K^{\ddagger}$ is the equilibrium constant between the reactants and the transition state for the reaction (either equation 2 or 4). The change in relative reaction rate due to an additive (X) at constant temperature and pressure can therefore be expressed as:

$$\left(\frac{\partial \ln k_{agg}}{\partial m_x}\right)_{T,P,m_P} = \left(\frac{\partial \ln K^{\ddagger}}{\partial m_x}\right)_{T,P,m_P} \qquad (6)$$

where m_x is the molality of additive. Using the Wyman linkage relation, the above expression can be written in terms of the extent of binding of the additive to the protein species:

$$\begin{aligned}\left(\frac{\partial \ln k_{agg}}{\partial m_x}\right)_{T,P,m_P} &= \left(\frac{\partial \ln a_x}{\partial m_x}\right)_{T,P,m_P}\left(\frac{\partial \ln k_{agg}}{\partial \ln a_x}\right)_{T,P,m_P} \\ &= \left(\frac{\partial \ln a_x}{\partial m_x}\right)_{T,P,m_P}(\Gamma^{\ddagger}_{XP} - \Gamma^{R}_{XP})\end{aligned} \qquad (7), (8)$$

where a_x is the thermodynamic activity of additive, and each Γ is a preferential binding coefficient. Wyman Jr., J. Adv. Protein Chem. 1964, 19, 223-286; Timasheff, S. N. PNAS 2002, 99, 9721-9726; Baynes, B. M.; Trout, B. L. J. Phys. Chem. B 2003, submitted for publication. $\Gamma^{\ddagger}_{PX}$ is the number of additive molecules bound to the transition state of equation 2 or 4, and Γ^{R}_{PX} is the number of additive molecules bound to the reactant in the same equation. Since $(\partial \ln a_X / \partial m_X)_{T,P,mp}$ is positive, equation 8 shows that in order for an additive to decrease the rate of aggregation, the additive must bind less to the transition state than to the reactant, making $\Gamma^{\ddagger}_{XP} - \Gamma^{R}_{XP}$ negative.

Attenuation of protein aggregation

In the pharmaceutical industry today, a refolding buffer additive used to increase the yield of active protein is the amino acid L-arginine. Arginine can increase the yield of active protein by an order of magnitude or more and has been successfully used to refold proteins such as tPA, interferon γ, lysozyme, carbonic anhydrase B, factor XIII, and antibodies. Rudolph, R.; Fischer, S.; Mattes, R. "German Patent DE 3 537 708, Process for the Activation of tPA after Expression in Prokaryotic Cells", **1985**; Arora, D.; Khanna, N. J. Biotechnol. **1996**, 52, 127-133; Armstrong, N.; de Lencastre, A.; Gouaux, E. Protein Sci. **1999**, 8, 1475-1483; Rinas, U.; Risse, B.; Jaenicke, R.; Abel, K.-J., Zettleneissl, G. Biol. Chem. Hoppe-Seyler **1990**, 371, 49-56; Buchner, J.; Rudolph, R. Biotechnology **1991**, 9, 157-162. Arginine has been shown to increase the yield of renatured protein by decreasing the rate of aggregation. Hevehan, D. L.; Clark, E. D. B. Biotechnol. Bioeng. **1997**, 54, 221-230. It is presently disclosed that arginine has a critical combination of two simple factors that enable it to prevent aggregation during folding. These factors include size and binding.

1. **Size.** Arginine is a much larger molecule than water, the primary solvent.

2. **Binding**. Protein molecules in isolation do not have a significant preference to be solvated by either arginine or water.

The mechanism by which the factors above affect aggregation is shown schematically in Figure 1. As the protein molecules diffuse toward each other, the size property ensures that a region of preferential hydration will form between the protein molecules because water but not the additive can fit in the gap (the oval in the transition state $A_2^{\ddagger}$ of Figure 1). This is analogous to "osmotic stress" effects on the equilibrium between two macromolecular conformations where one conformation has a crevice that water can enter but an additive cannot. Parsegian, V. A.; Rand, R. P.; Rav, D. C. PNAS USA 2000, 97, 3987-3992. The binding property ensures that when there is no steric constraint due to such a gap, arginine and water can solvate the protein equally well. This means that the region of preferential hydration shown in Figure 1 is the only contribution to the preferential binding coefficients of the additive with the protein in any of the three states shown (U + U, $A_2^{\ddagger}$, A_2). Because the transition state is preferentially hydrated, $\Gamma^{\ddagger}_{XP}$ is negative. Therefore the quantity $\Gamma^{\ddagger}_{XP} - \Gamma^{R}_{XP}$ is negative and aggregation is slowed. Any additive that has these two properties will deter aggregation during folding or in any other situation where a bimolecular step is rate limiting.

The size and binding properties are both necessary for prevention of aggregation. Molecules that meet the size criterion but not the binding criterion will either accelerate aggregation (such as "crowders" like dextran) or be denaturants (such as guanidinium chloride) and therefore have other undesirable effects on protein stability. Linder, R.; Ralston, G. Biophys. Chem. **1995**, 57; 15-25; Orsini, G.; Goldberg, M. E. J. Biol. Chem. **1978**, 253, 3453-3458; Jasuja, R. "B-114R Protein Drugs: Manufacturing Technologies", Technical Report, Business Communications Company, Inc., **2000**. A molecule that does not meet the size criterion but meets the binding criterion will have almost no effect on aggregation.

The two properties above differentiate molecules that may have advantageous effects on aggregation via the mechanism above from those that may not. It is believed that there are many molecules that have not been used as additives which have both of the above properties. Since these properties are presently disclosed, arginine was not selected with them in mind, implying that another yet untested molecule may exemplify the properties to a larger extent and have superior aggregation preventing characteristics. As non-limiting examples, some molecules with the two properties above that may prevent aggregation via a similar mechanism include:

- Citrulline
- Arginine or citrulline derivatives with a longer or shorter methylene linker between the amino acid backbone and guanidino or urea group (Figure 2).
- Arginine or citrulline derivatives where the amino acid backbone group is replaced by another large functional group which does not bind to proteins. (For example, 2-

guanidino acetic acid, 3-guanidino propanoic acid, 4-guanidino butyric acid, 5-guanidino pentanoic acid, etc.)

- Molecules that are not randomly orientated in solution near proteins. Such molecules can be constructed by covalently attaching a molecule which stabilizes proteins against unfolding with a molecule that destabilizes proteins against unfolding. Examples of novel molecules designed based on this idea are shown in Figure 3. A partial list of molecules that are known to stabilize and destabilize proteins against unfolding are shown in Table 1.

Table 1.

Protein Stabilizer	Protein Destabilizer
Sugars (e.g. glucose, sucrose, trehalose)	Urea
Polyols (e.g. sorbitol, mannitol)	Guanidinium chloride
Dextran	Detergents (e.g. sodium dodecyl sulfate, Tris)
Kosmotropes	Chaotropes
Glycine, glycine betaine	

Compounds of the present invention

Based on the studies described in the previous section, compounds of the present invention may be prepared by functionalizing a molecule that does not bind to a protein with at least one protein binding group. In other words, compounds of the present invention possess a non protein bonding moiety and a protein binding group. Molecules that do not bind to proteins include but are not limited to osmolytes and kosmotropes, such as glycerol, glycine betaine, dendrimers, and trimethyl amine N-oxide. Other such molecules are known to those skilled in the art.

A protein-binding group is a molecule or functional group that binds to some proteins. Many molecules that fall in this class are, for example, denaturants or surfactants. Some non-limiting examples of protein-binding molecules are: the guanidinium ion, urea, amino acids (such as arginine, lysine, aspartate, glutamate), sodium dodecyl sulfate, tweens (polysorbate), poloxamers, and ions (such as citrate and acetate). A group or molecule does not need to bind to all proteins to be classified as a "protein-binding group;" rather, it merely needs to bind to some proteins. The concepts of "binding" and groups or molecules that bind to proteins are well-known to those skilled in the art.

The net effect of functionalizing a non-binder with a protein-binding group will be to move the protein preferential binding coefficient toward zero. Molecules that are large, but have a protein preferential binding coefficient near zero, have the properties that they prevent aggregation but do not destabilize native protein molecules. Thus, these molecules are useful as anti-aggregation additives.

Statistical model approach for stabilizing proteins towards aggregation

Additives perturb the chemical potential of the protein system by associating either more strongly or more weakly with the protein than water. This phenomenon, called "preferential binding," is of great interest because it governs the physical and chemical properties of proteins. Timasheff; S. N. *Adv. Protein Chem.* **1998**, 51, 355-431.

When an additive (X) is added to an aqueous protein solution, it alters the chemical potential of the protein (μ_P) via the following relationship:

$$\begin{aligned}\Delta\mu_P^{tr} &= \int_0^{m_X}\left(\frac{\partial\mu_P}{\partial m_X}\right)_{m_P} dm_X \\ &= -\int_0^{m_X}\left(\frac{\partial\mu_X}{\partial m_X}\right)_{m_P}\left(\frac{\partial m_X}{\partial m_P}\right)_{\mu_X} dm_X\end{aligned} \qquad (9),\,(10)$$

where $\Delta\mu_P$ is the transfer free energy of the protein from pure water into the mixed solvent system, *m* is molality, and subscripts *X* and *P* identify the additive and protein respectively. Lee, J. C.; Timasheff, S. *N. J. Biol. Chem.* **1981,** 256, 7193-7201. Two partial derivatives appear in equation 10. The first captures the dependence of the additive chemical potential on additive molality and can be evaluated by experiments on a binary mixture of additive and water ($m_P \rightarrow 0$). The second partial derivative is the "preferential binding coefficient;" Γ_{XP}:

$$\Gamma_{XP} \equiv \left(\frac{\partial m_X}{\partial m_P}\right)_{\mu_X} \qquad (11)$$

The preferential binding coefficient is a way in which binding can be defined thermodynamically. It is also particularly useful when binding is weak. The preferential binding coefficient is a measure of the excess number of additive molecules in the domain of the protein per protein molecule (Figure 4). The connection between the thermodynamic definition (equation 11) and the intuitive notion of binding (local excess number of molecules) comes from statistical mechanics; where it can be shown that:

$$\Gamma_{XP} = \left\langle n_X^{II} - n_W^{II}\left(\frac{n_X^{I}}{n_W^{I}}\right)\right\rangle \tag{12}$$

In the above equation, n denotes the number of a specific type of molecule (subscript X for the additive and subscript W for water) in a certain domain (superscript I for a bulk volume outside of the vicinity of the protein and superscript II for a volume in the protein vicinity), and angle brackets denote an ensemble average. Kirkwood, J. G.; Goldberg, R. J. *J. Chem. Plays.* **1950,** 18, 54-57; Schellman, J. A. *Biopolymers* **1978,** *17,* 1305-1322. Note that Γ_{XP} is independent of the choice of the boundary between the domains, as long as the boundary is far enough from the protein.

If the additive concentration is higher in the vicinity of the protein than in the bulk, Γ_{XP} is greater than zero, and μp is lower in the presence of the additive than in its absence. Denaturants such as urea and guanidinium chloride exhibit this type of binding behavior. The reverse is true for sugars, such as trehalose. In trehalose solutions, there is generally a deficiency of trehalose and an excess of water in the vicinity of the protein. For this "preferential hydration" case, Γ_{XP} is less than zero, and μp is higher in the presence of the additive.

Timasheff pioneered the use of high-precision densitometry to measure preferential binding coefficients for protein-cosolvent systems. Lee, J. C.; Timasheff, S. *N. J. Biol. Chem.* **1981,** 256, 7193-7201; Lee; I. C.; Timasheff; S. N. *Biochemistry* **1974,** *13.* 257-265; Gekko, K.; Timasheff, S. N. *Biochemistry* **1981,** *20.* 4667-4676; Gekko, K.; Timasheff, S. N. *Biochemistry* **1981,** *20,* 4677-4686. More recently, differential scanning calorimetry (DSC) and vapor pressure osmometry (VPO) have been used to the same end. Poklar, N.; Petrovcic. N.; Oblak, M.; Vesnaver; G. *Protein Sci.* **1999,** 8, 832-840; Courtenay, E. S.: Capp, M. W.; Anderson; C. F.; Record Jr., 11. *T. Biochemistry* **2000,** *39,* 4455-4471. Preferential binding coefficients are rigorous thermodynamic quantities and are related to virial coefficients, activity coefficients, and free energies via standard thermodynamic relations for multi-component solutions. Casassa. E. F.; Eisenberg, *H. Adv. Protein Chem.* **1964,** *19,* 287-395.

Experimental studies by the above methods have led to some generalizations about preferential binding coefficients:

1. Γ_{XP} may be positive or negative, indicating that interactions of the protein and additive are favorable or unfavorable, respectively.

2. Γ_{XP} is proportional to additive molality at low concentration of additive (often as high as *mx* ~ 1 m and higher). Courtenay, E. S.: Capp, M. W.; Anderson; C. F.; Record Jr., 11. *T. Biochemistry* **2000,** *39,* 4455-4471; Greene Jr., R. F.; Pace. C. *N. J. Biol.Chem.* **1974,** *249,* 5388-5393; Record Jr., M. T.; Zhang; W.; Anderson; C. F. *Adv. Protein Chem.* **1998,** *51,* 281-353.

3. Γ_{XP} is roughly proportional to the protein-solvent interfacial area. Lee, J. C.; Timasheff, S. *N. J. Biol. Chem.* **1981,** 256, 7193-7201.

The second generalization above, together with the fact that many binary mixtures of additive and water (*mp* ⟶ 0) are nearly ideal at low concentration of additive, leads to a useful simplification of equation 10:

$$\begin{aligned}\Delta\mu_P^{tr} &= -\int_0^{m_X}\left(\frac{\partial RT\ln m_X}{\partial m_X}\right)_{m_P}\left(\frac{\Gamma_{XP}}{m_X}\right)m_X dm_X \\ &= -RT\left(\frac{\Gamma_{XP}}{m_X}\right)\int_0^{m_X} dm_X \qquad (13),(14),(15) \\ &= -RT\,\Gamma_{XP}\end{aligned}$$

Equation 15 provides a simple and convenient link between preferential binding coefficients and free energies. This relation leads to the useful rule that when Γ_{XP} is proportional to *mx*, for each additive molecule that preferentially interacts with the protein, the protein's free energy is reduced by approximately 0.6 kcal/mol at 25°C. The simplicity of this relation is a natural result of the close relationship between Γ_{XP} and a second virial coefficient.

To be able to predict preferential binding coefficients and understand their origins, the above thermodynamic framework and general observations must be augmented by a mechanistic model. Several such models have been presented in the literature, including models based on the binding polynomial or statistical mechanical partition function, solvent-additive exchange at defined sites, additive partitioning between the local and bulk domains, and group contribution methods for estimating transfer free energies.

The most general model of additive binding hitherto presented comes from considering an equilibrium of all possible protein-additive complexes, from which it can be shown that:

$$\Delta\mu_P^{tr} = -RT\ln(1+\sum_i\sum_j K_{ij}m_W^i m_X^j) \qquad (16)$$

where K_{ij} is the equilibrium constant for a reaction of a protein molecule, *i* molecules of water, and *j* molecules of additive into a complex. Wyman, J.; Gill; S. J. *Binding and Linkage: Functional Chemistry of Biological Macromolecules:* University Science Books: 1990. While this model is completely general, its utility is limited because it is not possible to determine experimentally the many K_{ij} parameters present in equation 16.

Schellman's *site exchange model*, provides a way to simplify this general expression to a form containing a single parameter. Schellman, J. A. *Biopolymers* **1978,** *17,* 1305-1322. This model treats binding as a family of protein-solvent exchange reactions such as:

$$\mathrm{P\cdot W_i + X \rightarrow P\cdot X + iW} \qquad (17)$$

where P is the protein, W is water, X is cosolvent; and i is the exchange stoichiometry. The simplification requires the assumptions that 1:1 exchange reactions (i = 1) occur on a fixed number of identical, independent sites and that the sites are far from saturation with additive (i.e. the apparent dissociation equilibrium constant for each site is well above the additive concentration). The number of sites, *n*, is approximated by the number of water molecules present in a monolayer around the protein. These simplifications reduce equation 16 to:

$$\Delta\mu_P^{tr} = -nRT\langle K\rangle m_X \qquad (18)$$

where $\langle K\rangle$ is the average equilibrium constant of binding at a single site. The single parameter $\langle K\rangle$ can then be determined from an experimental measurement of Γ_{XP}. When equation 15 holds, the relation between $\langle K\rangle$ and Γ_{XP} is simply:

$$\langle K\rangle = \Gamma_{XP} / nm_X \qquad (19)$$

Values of $\langle K \rangle$ for different proteins in this linear regime are roughly equal. Schellman, J. A. *Biophys. Chem.* **2002,** *96.* 91-101. $\langle K \rangle$ cannot, however, be determined without knowledge of Γ_{XP} or other free energy data on the particular additive system of interest. In fact, one can say that $\langle K \rangle$ is defined by Γ_{XP}.

Another model that recasts preferential binding coefficient data in terms of a single model parameter is the *local-bulk domain model* developed by Courtenay et al. Courtenay, E. S.: Capp, M. W.; Anderson; C. F.; Record Jr., 11. *T. Biochemistry* **2000,** *39,* 4455-4471. The parameter in this model is the partition coefficient K_p, relating the number of water molecules and additive molecules in the local and bulk domains via:

$$K_P = \frac{n_X^{II} / n_W^{II}}{n_X^{I} / n_W^{I}} \tag{20}$$

Similar to the site exchange model, the convention used in this model is that the local domain consists of a monolayer of water and enough additive to obtain the experimentally observed Γ_{XP}. Note that because the absolute occupancy of water and additive in the local domain cannot be easily determined by experiment, the local-bulk domain model effectively defines *nw*. Like $\langle K \rangle$, values of K_p can be used to predict Γ_{XP} at other additive concentrations or for other proteins in the same additive, but predictions cannot be made in the absence of Γ_{XP} or free energy data on the same additive system.

Lastly, transfer free energy models, pioneered by Bolen's group, take a different approach. Liu, Y. F.; Bolen, D. W. *Biochemistry* **1995,** *34,* 12884-12891. These models conceptually divide whole proteins into groups such as the amino acid side chains and the protein backbone and model the transfer free energy of the whole protein as a sum of the transfer free energy of the groups it comprises, via:

$$\Delta\mu_P^{tr} = \sum_i \alpha_i \Delta g_i^{tr} \tag{21}$$

where Δg_i is the transfer free energy of the model group and α_i is the solvent accessible area of the group in the whole protein, normalized to the solvent accessible area of the model compound. Tanford, C. *J. Am. Chem. Soc.* **1964,** *86,* 2050-2059. The overall $\Delta\mu^{tr}p$ can then be predicted for any system of known structure. In the context of the previously described models, the transfer free energy model can be thought of as a linearized binding model where each surface group or amino acid in the protein represents a different type of independent binding site, and the binding constants for those sites are determined by experiments on model compounds, such as free amino acids or cyclic di-amino acid compounds. Predictions made by transfer free energy models have met with mixed success. A linear group contribution model (equation 21) may be too simple to capture all of the important contributions to $\Delta\mu^{tr}p$. Bolen, D. W. Protein Stabilizaiton by Naturally Occurring Osmolytes. In *Protein Structure, Stability, and Folding;* Humana Press: 2001.

While the above models have helped in the understanding of the phenomenon of preferential binding, they generally incorporate strong assumptions, and they necessitate the use of experimental data on highly analogous systems in order to determine model parameters and make predictions. Thus, their uses as predictive tools and as tools to gain insight into specific systems are limited.

One aspect of the present invention relates to a predictive, molecular-level approach for the study of preferential binding based on all-atom, statistical mechanical models that use no adjustable parameters. To date, statistical mechanical models of preferential binding have only been developed for interactions of ions with charged cylinders and for interactions of two-dimensional, "hard circles" with a linear interface, both far too simple to be generally applied to protein-additive systems. Anderson; C. F.; Record Jr., M. *T. J. Phys. Chem.* **1993,** *97,* 7116-7126; Mills, P.; Anderson, C. F.; Record Jr., M. *T. J. Phys. Chem.* **1986,** *90,* 6541-6548; Tang. K. E. S.: Bloomfield, V. A. *Biophys. J.* **2002,** *82.* 2876-2991. Other explicit mixed solvent simulations of proteins and amino acids have been performed, but these studies did not compute thermodynamic quantities related to preferential binding. Zou, Q.; Bennion. B. J.; Daggett, V.; Murphy, K. P. *J. Am. Chem. Soc.* **2002,** *124,* 1192-1202; Bennion, B. J.; Daggett, V. *PNAS* **2003,** *100,* 5142-5147; Tirado-Rives, J.; Orozco, M.; Jorgensen, W. L. *Biochemistry* **1997,** *36,* 7313-7329; Alonso, D. O. V.; Daggett, V. *J. Mol. Biol.* **1995,** *247,* 501-520; Caflisch. A.; Karplus, *XI. Structt. Fold. Des.* **1999,** *7,* 477-488. In the present invention, the number of "bound" molecules are defined in a thermodynamically consistent way and do not *a priori* incorporate any information about "binding sites." The use of this approach for the computation of preferential binding coefficients was validated in two systems by comparison with experimental data from the literature. Additionally, the molecular-level detail of the approach provides new insights into the following issues:

1. The changes in solvent and additive concentration as a function of distance from the protein surface.
2. A precise definition of the "local domain" (Figure 4).
3. The differences in preferential binding or apparent binding equilibrium constant at different locations on the protein-solvent interface.

The success of this method in modeling preferential binding indicates that it captures the important underlying physics of protein-additive-water systems and that the difficulty in quantitative prediction to date can be surmounted by explicitly incorporating the complex protein-solvent and solvent-solvent interactions.

A Molecular-Level Approach to Computing Preferential Binding

One aspect of the present invention relates to the use of explicit atomic interaction potentials (force fields), such as Lennard-Jones, Coulombic, spring, and torsion interactions, with pre-fit coefficients. Brooks; B. R.; Bruccoleri; R. E.; Olafson, B. D.; States, D. J.; Swaminathan, W.: Karplus, M. *J. Comp. Chem.* **1983,** *4,* 187-217; Ha; S. N.; Giammona; A.: Field, M.; Brady, J. W. *Carbohydrate Res.* **1988,** *180,* 207-221. Thermodynamic properties, such as preferential binding coefficients, are computed by averaging in the time domain via molecular dynamics (MD). A snapshot from a dynamic simulation of RNase T1 in a urea solution is shown in Figure 5, which was generated with VMD. Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996,** *14,* 33-38. The results of the simulations contain all of the information needed to extract thermodynamic properties, such as Γ_{XP}.

Molecular dynamics uses Newton's second law of motion, that acceleration is the quotient of force and mass, to compute the positions of each atom in the system as a function of time. To do this, an energy model, sometimes called a "force field," that can be used to compute the net force on any atom in any configuration is employed.

During the MD run, the positions of each atom are recorded at fixed intervals in time. These "snapshots" form an ensemble of configurations which can then be used to compute thermodynamic properties, such as Γ_{XP}.

Importantly, this method of computing Γ_{XP} does not introduce any adjustable parameters to model preferential binding or any other aspect of a system containing a protein and solvent-additive components. All of parameters required by the MD method for energy computations are determined independently of this particular modeling objective, and in fact have been shown to be generally applicable to biological systems. Karplus, M., McCammon, J. A. *Nature. Struct. Biol.* **2002,** *9,* 646-652. Thus, the method developed here could be used to estimate Γ_{XP} and $\Delta\mu^{tr}p$ in systems where no experimental data is available. It therefore facilitates the study of preferential binding when direct experimental study is difficult, such as at transition state configurations or at marginally stable states of proteins. Furthermore, it yields detailed, local, molecular-level insight into the system studied.

Another benefit of this approach is that when equation 15 holds (such as for urea and glycerol), the protein transfer free energy ($\Delta\mu^{tr}p$) can be calculated from a single Γ_{XP} simulation. Traditional free energy calculation methods such as thermodynamic integration require 15-20 trajectories, which is computationally difficult for protein systems of this size. Bash, P. A.; Singh, U. C.: Langridge, R..; Kollman. *P. A. Science 87, 236,* 564-569; Kollman, *P. Chem. Rev.* **1993,** *93,* 2395-2417.

Preferential Binding Coefficients of Constituent Groups

Because proteins have a range of different functional groups in different orientations on their surfaces, the concentrations of solvents and additives near different patches on the protein's surface may be different. For example, the vicinity of a hydrophobic patch on the protein may have a lower concentration of water and a higher concentration of additive than in the vicinity of a hydrophilic patch. Preferential binding experiments capture only the average effect arising from all of the interactions over the entire protein-solvent interface; however, molecular simulations allow more detailed analyses of the local contributions to preferential binding coefficients.

A protein can be thought of as a set of non-overlapping constituent groups, each of which has its own preferential binding coefficient defined by the composition of the solvent in its immediate vicinity. Tanford, C. *J. Am. Chem. Soc.* **1964,** *86,* 2050-2059. Similar to group contribution methods for computing transfer free energies, one possible group definition is that each type of amino acid side chain (up to 20) and the amino acid backbone are distinct groups.

To compute a preferential binding coefficient for a constituent group, the solvent molecules in the local domain are assigned only to the nearest group (i), and the "group preferential binding coefficients" ($\Gamma_{XP,}$ i) can be defined as:

$$\Gamma_{XP,i} = \left\langle n^{II}_{X,i} - n^{II}_{W,i}\left(\frac{n^{I}_{X}}{n^{I}_{W}}\right)\right\rangle \tag{22}$$

where $n^{II}_{x,i}$ and $n^{II}_{w,i}$ are the number of additive and water molecules in the local domain that are nearest to group *i*. If each additive molecule in the local domain is assigned to a group, the overall preferential binding coefficient is simply the sum of all of the group preferential binding coefficients:

$$\Gamma_{XP} = \sum \Gamma_{XP}, i \tag{23}$$

The group preferential binding coefficients decompose the effect of each small subset of the protein on the overall preferential binding coefficient. This is analogous to the group contribution models for transfer free energy except that the parameters are extracted from a simulation of an entire protein instead of experiments on model compounds.

Minimum Simulation Time

Sufficient sampling of position-space configurations in time is required for the accurate calculation of Γ_{XP} via equation 11. Assuming that the average protein solution structure is close to that of the initial (crystal) structure and that water molecules sample position space rapidly because of their high density, the most important time scale to be captured is that of the additives sampling position space. One way to estimate this time is that it must be much larger than the average time between additive-additive contacts.

An estimate of the time between contacts can be obtained as:

$$t_{contact} \approx \frac{1}{12D}\left(\frac{V_{soln}}{n_X}\right)^{\frac{2}{3}} \tag{24}$$

where D is the additive diffusivity, V_{solv} is the solvent volume, and nx is the number of additive molecules. For the simulations performed here, the solvent is mostly water, so equation 24 can be further simplified to yield:

$$t_{contact} = \frac{1}{12D}\left(\frac{1}{N_A \rho_W m_X}\right)^{\frac{2}{3}} \tag{25}$$

where N_A is Avogadro's number and ρw is the density of water in kg/m^3. For a 1 m additive in water system with a additive diffusivity of 2x10^{-9} m^2/s (a lower bound on the diffusivities of the additives studied here), $t_{contact}$ is about 30 ps. Thus, nanosecond trajectories will be required for good sampling of additive position space. Importantly, this time increases as the additive concentration decreases, implying that there is a minimum concentration that can be studied with any given amount of computational resources.

Radial Distribution Functions of Water and Additives

The radial distribution functions of water, urea, and glycerol were computed for all three simulations as described in the Exemplification section and are shown in Figure 6.

At very short distances, r < 0.6Å for water and r < 1.0Å for glycerol and urea, regions of total solvent and additive exclusion due to very strong van der Waals repulsion can be seen. The size of these "totally excluded" regions is much smaller than one would expect based on the apparent van der Waals radii of the solvent and additive molecules alone (for example, $r \approx$ 1.5Å for water and 2.2Å for urea), indicating that electrostatic attractive forces play an important role in solvation even at these distances. Schellman, J. *A. Biophys. J.* **2003,** *85*. 108-125. After the regions of total exclusion, strong first coordination shells of these three molecules can be clearly seen. The peaks of the first coordination shells become more distant from the protein as the size of the molecules they correspond to increases. Significantly smaller second coordination shell peaks are also visible for urea solvating RNase T1 and glycerol solvating RNase A. At distances greater than 6-7Å from the protein, solvation shells cannot be discerned, and the number densities of water, urea, and glycerol reach their bulk values.

In the simulations of RNase T1 in glycerol and urea solutions, the radial distribution functions for glycerol and urea are quite different. The maximum value of $gx(r)$ for urea is over 4.5, while that for glycerol is about 2.5. The difference in these maximum values, while significant, is not sufficient to say that the number of urea molecules coordinated to the protein (n_x) is higher than the number of glycerol molecules coordinated, this can only be done by integrating each $gx(r)$ function appropriately via equation 31.

The radial distribution functions for both water and glycerol are similar in the simulations of RNase A and RNase T1 in glycerol solution, despite the fact that the proteins and the pHs of the solutions are different. Given that the proteins are of similar size, this observation is consistent with the fact that the values of Γ_{XP} for the two solutions are close.

Preferential Binding Coefficients

The radial distribution functions in Figure 6 suggest that r_* in the range of 6-8Å is an appropriate choice of boundary between the local and bulk domains. The error in Γ_{xp} introduced by a particular choice of the boundary distance, r_*, can be estimated by plotting the apparent preferential binding coefficient (Γ_{xp}) versus r_* (Figure 7). Γ_{xp} depends very strongly on r_* in the first solvation shell (r = 0 - 4Å) and weakly on r_* in the second solvation shell (r = 4 - 6Å). In the range r = 6 - 8Å, the dependence of Γ_{xp} on r_* is small (±0.5), and is less than the statistical error in Γ_{xp} (shown in Table 2, explained below). Therefore, a cutoff distance of 6Å, or about two solvation shells, is sufficiently large to minimize systematic error in Γ_{xp} caused by the choice of r_*. If only a single solvation shell were considered (r_* ~ 3.5 - 4Å), a systematic error in Γ_{xp} of approximately 0.5 - 1 molecules would be introduced as a result of neglect of the second solvation shell.

The preferential binding coefficient, Γ_{xp}, was computed via equation 11 using r_* = 6Å as the boundary between the local and bulk domains. A confidence interval for this ensemble average was computed as described in the Exemplification section. The binding coefficients and their statistical uncertainties are shown in Table 2.

Table 2. Preferential binding coefficients computed from MD simulations and compared with available experimental data at similar additive concentrations.

System	m_{bulk}	Simulation Γ_{XP}	Experimental Γ_{XP}
Urea / Rnase T1	1.10 m	5.2 ± 1.0	6.4[a]
Glycerol / Rnase T1	1.07 m	-1.6 ± 0.8	

Glycerol / Rnase A	0.91 m	-0.9 ± 1.0	-1.7 ± 0.8[b]

[a] Lin, T. Y.; Timasheff, S. N. *Biochemistry* **94,** *33,* 12695-12701.

[b] Gekko, K.; Timasheff, S. N. *Biochemistry* **1981,** *20.* 4667-4676.

A wide range of behavior (positive and negative preferential binding coefficients) can be modeled without the use of adjustable parameters. The confidence intervals on Γ_{xp}(MD) are an estimate of the statistical error resulting from the use of a finite trajectory. For easier comparison, the experimental values of Γ_{xp} reported above were interpolated to m_{bulk} from data sets spanning the molality of interest.

Experimental values from the literature were available for two out of three of these protein-additive systems, and the computed values of Γ_{xp} agree quite favorably with these values. The fact that this occurs for both positive and negative values of Γ_{xp} without the use of any adjustable parameters is very encouraging. For an additive that obeys equation 15, the confidence intervals of ±1.0 in Γ_{xp} represents a confidence limit in the transfer free energy of about 0.6 kcal/mol, which is a typical value for free energies calculated via this type of molecular simulation. Achievement of this level of accuracy despite the fact that structural fluctuations in the native state ensemble of proteins have been observed on much longer time scales than the time scale of the simulations performed here suggests that solvent dynamics are more important than protein structural dynamics in determining Γ_{xp}. Duan, Y.; Kollman, P. A. *Science* **1998,** *282,* 740-744.

$\Gamma_{xp}(t)$ probability density functions for the simulations of RNase T1 in urea and glycerol solution are shown in Figure 8. The range of instantaneous values of the preferential binding coefficient, $\Gamma_{xp}(t)$, is quite large relative to the absolute values of Γ_{xp}. $\Gamma_{xp}(t)$ values in excess of Γ_{xp} ± 15 are observed. The breadths of these distributions are related to the size of the interface between the local and bulk domains and indicate the importance of sampling a large number of solvent configurations to obtain the macroscopic, averaged Γ_{xp} (equation 27).

The Relation between Solvent Accessible Area and the Number of Molecules in the Local Domain

The solvent accessible areas of whole proteins (SAA) and constituent groups (SAA_i) in crystal structures have been used extensively in analyzing proteins. SAA and SAA_i are essentially simple ways of measuring water coordination numbers. In models developed to date, SAA or SAA$_i$, has been used to estimate n_w or $n_{w,i}$ by assuming that the local domain is a inonolayer of water and each water molecule occupies approximately 10Å^2 of the solvent accessible area. Since the present invention introduces a new notion of the local domain, it is worthwhile to see what relationships exist between SAA$_i$ and the coordination numbers $n_{w,i}$ and $n_{x,i}$ that utilize this definition.

A scatter plot of the solvent accessible area of a set of constituent groups (amino acid side chains and the protein backbone) versus the number of water molecules in the local domain for three different simulations is shown in Figure 9. Solvent accessible area was calculated analytically in CHARMM (based on Richmond's method) using a 1.4Å probe. Richmond, T. J. *J. Mol. Biol.* **1984,** *178,* 63-89. There is a strong, linear correlation of these variables with slope 4.2 Å^2/molecule and correlation coefficient 0.96. Similarly strong correlations are seen for SAA$_i$ with $n_{x,i}$ in individual simulations. A summary of proportionality constants and correlation coefficients for these relationships is shown in Table 3. If the time average SAA$_i$ from each dynamics simulation is used instead of the crystal structure SAA$_i$ values, the correlation coefficients increase slightly. Because the time average solvent accessible areas arc higher than those in the crystal structure, the proportionality constants shown in Table 3 also increase.

Table 3. Relationships between solvent accessible area in each protein crystal structure and number of solvent molecules in the local domain for different protein-additive systems. r^2 symbolizes the correlation coefficient.

Species (i)	Protein	Avg. Protein SAA/n_i^{II} (Å^2/molecule)	r^2
Water	RNase A/T1	4.2	0.96
0.91 m Glycerol	RNase A	290	0.96
1.07 m Glycerol	RNase T1	230	0.93
1.10 m Glycerol	RNase T1	170	0.98

Constituent Group Preferential Binding Coefficients

The constituent group preferential binding coefficients were calculated for each simulation as described in the Exemplification section and are shown in Figures 10 - 13 as the number of water and additive molecules coordinated to each constituent group. In each figure, a line at the bulk solution composition is also plotted, enabling a quick determination of the composition of the solvent in the vicinity of a constituent group compared to the bulk solvent. The statistical uncertainties in the values of $n^{II}{}_{w,i}$ and $n^{II}{}_{x,i}$ (and consequently $\Gamma_{xp,i}$) are high. Because of these uncertainties, we will not report specific values of the group preferential binding coefficients, but rather classify them into broad categories based on their statistical likelihood of being either positive, negative, or zero/ indeterminate.

The average number of water and glycerol molecules coordinated to each of the 15 serine residues in RNase T1 are shown in Figure 10. A wide range of binding behavior can be seen among the serine residues, all of which have a good degree of solvent exposure. Ser 17, 35, and 72 fall above the bulk concentration line and have positive preferential binding coefficients, Ser 63 falls below the line and has a negative preferential binding coefficient, and the preferential binding coefficients of the remaining 11 serine residues are not statistically different from zero. The wide range of local concentrations in the vicinities of these serine residues indicates that developing a group contribution method to estimate Γ_{xp} or $\Delta\mu^{tr}p$ based on primary sequence information and solvent accessibility ($n^{II}{}_{w,i}$) alone may be difficult. In addition to the type of amino acids present at the protein-solvent interface, other effects such as specific combinations of residues and secondary or tertiary structure must be important in determining water and additive binding behavior. These factors probably contribute to the range of local concentrations seen in Figure 10. For example, Ser35 and Ser72 are proximal to each other and several Gly and Tyr side chains (Gly 34, 70, 71, and Tyr 68), which tend to have positive preferential binding coefficients in glycerol (Figure 12). This may be the reason that the group preferential binding coefficients for these residues are higher than those of the other serine residues.

The preferential binding behavior of urea and glycerol, with each type of amino acid in RNase T1 and the protein backbone are shown in Figures 11 and 12. In urea solution, the protein backbone and Ser as well as the hydrophobic amino acid side chains of Cys, Gly, Len, Phe, Pro, Tyr, and Val all preferentially bind urea, while the hydrophilic Asp preferentially binds water. In glycerol solution, only Tyr and Gly preferentially bind glycerol, and Asp and Glu preferentially bind water. Qualitatively, the binding behavior of the amino acid side chains of RNase T1 follow a hydrophobic series, with the hydrophobic side chains tending to bind more additive and the hydrophilic ones tending to bind more water.

The binding behavior of glycerol and water with the amino acid side chains and backbone in RNase A, shown in Figure 13, is significantly different than the binding behavior of these solvent components with the same constituent groups in RNase T1. (Note that the protonation states of Asp, Glu, and His are different in the two simulations.) The amino acid backbone, which occupies a large fraction of the protein-solvent interface as indicated by its high value of $n^{II}_{w,i}$, has a binding coefficient near zero in RNase T1 and a significant negative binding coefficient in RNase A. More strikingly, Tyr in RNase T1 preferentially binds glycerol whereas Tyr in RNase A preferentially binds water. This is likely because the six Tyr residues in RNase A are at or near the solvent interface (a more hydrophilic region) whereas the nine in RNase T1 are mostly buried (a more hydrophobic region). This difference in solvent exposure is evident from the crystal structures of the proteins but also can be discerned by comparing the water coordination numbers for Tyr in the two proteins: $n^{II}_{w,i}$ for Tyr in RNase A is higher than in RNase T1, even though there are 50% more Tyr residues in RNase T1.

Based on the above observations, some generalizations about the effects that these additives have on protein folding equilibria can be postulated, the validity of which must be confirmed via future studies. In urea solution, most of the constituent groups in RNase T1 either preferentially bind urea or are indifferent to urea and water. Asp, which is found on the surface of RNase T1, is the only constituent group that is significantly below the bulk concentration line in Figure 11 and therefore preferentially binds water over urea. Since the amino acids that compose the core of RNase T1 and are exposed upon unfolding preferentially bind urea, this pattern suggests that the preferential binding coefficient or urea with unfolded RNase T1 is higher than that with native RNase T1. This is thermodynamically consistent with urea's well-known ability as a denaturant. Inversely, in glycerol solution, almost all of the constituent groups in RNase A and TI are neutral or preferentially bind water. This is consistent with the fact that glycerol binds less to the unfolded protein than the native state, and therefore is a protein stabilizer. Both of these generalizations are consistent with earlier work on model compounds. Bolen, D. W. Protein Stabilizaiton by Naturally Occurring Osmolytes. In *Protein Structure, Stability, and Folding;* Humana Press: 2001.

Exemplification

The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

Example 1

Molecular Simulations - Molecular dynamics was used to sample the phase space of proteins solvated by water and an additive. Version 28 of the CHARMM molecular dynamics package was used for all simulations. Brooks; B. R.; Bruccoleri; R. E.; Olafson, B. D.; States, D. J.; Swaminathan, W.: Karplus, M. *J. Comp. Chem.* **1983,** *4,* 187-217. The CHARMM force-field was used for the protein, and the TIP3P model [32] was used for water. Jorgensen, W. L.; Chandrasekhar. J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983,** *79,* 926-935. A force-field was constructed for glycerol using the standard CHARA-II\-1 geometries and partial charges for the atoms in a -CHOH- unit. Brooks; B. R.; Bruccoleri; R. E.; Olafson, B. D.; States, D. J.; Swaminathan, W.: Karplus, M. *J. Comp. Chem.* **1983,** *4,* 187-217; Ha; S. N.; Giammona; A.: Field, M.; Brady, J. W. *Carbohydrate Res.* **1988,** *180,* 207-221. Urea was assumed to be planar with bond lengths equal to the CHARMM standards and partial charges recomputed as done previously [33] but using the CHARMM van der Waals mixing rules in the objective function. Duffy. E. M.; Severance. D. L., Jorgensen, W.L. *Israel J. Chem.* **1993,** *33,* 323-330.

The structures of RNase A (PDB code: lfs3) and RNase T1 (PDB code: lygw) were obtained from the Protein Data Bank. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliand; G.; Bhat; T. N.; Weissig, H.; Shindyalov. I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000,** *28,* 235-242. In total; three simulations were performed: RNase A in lm glycerol (pH 3), RNase T1 in lm glycerol (pH 7), and RNase T1 in lm urea (pH 7). Details of each simulation are shown in Table 4. Each protein was solvated in a truncated octahedral box extending a minimum of 9A from the protein. The pH of each simulation was fixed by setting the protonation states of each ionizable side chain to the dominant form expected for each amino acid at the pH of interest. Arginine, cysteine, lysine, and tyrosine were protonated in all of the simulations. Aspartate, glutamate, and histidine were assumed to have pKa values of 3.4, 4.1, and 6.6, respectively; and were therefore protonated in the simulation at pH 3 and deprotonated at pH 7. Forsyth, W. R.;

Antosiewicz. J. hl.; Robertson, A. D. *Proteins* **2002,** *48,* 388-403; Edgecomb, S. P.; Murphy, K. *P. Proteins* **2002,** *49,* 1-6. Initial placement of water and additive molecules were random. Protein counterions were placed using SOLVATE 1.0. The system was first energy minimized at 0 K, next heated to 298.15 K, and then equilibrated for 1 nanosecond in the NTP ensemble at one atmosphere. For the computation of the properties of interest, two nanoseconds of dynamics were then run, during which statistics were computed from snapshots of the trajectory every picosecond.

Table 4. Details of four molecular dynamics (AID) simulations performed. *n*x is the number of additive molecules, n_w is the number of water molecules, and <*l*> is the average dimension of the primary unit cell (which varies during the run at constant pressure).

Additive	Protein	T (°C)	pH	n_x	n_w	<*l*> (Å)
Urea	RNase T1	25	7	90	4274	57.48
Glycerol	RNase T1	25	7	87	4582	59.24
Glycerol	RNase T1	25	3	90	5480	62.86

Example 2

Calculation of Preferential Binding Coefficients - The trajectories were then used to define the local and bulk regions and compute Γ_{xp} in the following manner. For the purpose of computing Γ_{xp} and other thermodynamic and structural parameters, each water and additive molecule was treated as a point at its center of mass. The distance of each of these points to the protein's van der Waals surface was computed, and then ρ*w*(*r*) and ρx(*r*), defined as the number densities of these points at a distance r from the protein, were computed. In all cases, the ρ(*r*) functions exhibited peaks and valleys characteristic of solvation shells in the range $0 < r < 6$Å. At distances in the range of 6-8Å and higher, such variations are no longer seen, and the local number density is defined as bulk number density, $\rho(\infty)$. Such a region far from the protein containing a spatially uniform concentration of water and additive must be present in the simulation cell in order to define the local and bulk regions and calculate Γ_{xp}.

The position of the boundary between the local and bulk domains, a distance of r_* away from the surface of the protein, was then determined by choosing the minimum distance at which no significant difference between ρ(r_*) and ρ(∞) was apparent for either water or additive. All solvent molecules whose centers of mass fell inside a distance of r_* from the protein's van der Waals surface were defined as belonging to the local domain (II), and all other solvent molecules were defined as belonging to the bulk domain (I). With these definitions of the domains, the instantaneous preferential binding coefficient, $\Gamma_{xp}(t)$, was computed as

$$\Gamma_{XP}(t) \equiv n_X^{II} - n_X^{I} \left(\frac{n_W^{II}}{n_W^{I}} \right) \qquad (26)$$

for each time point in each trajectory. The preferential binding coefficient, Γ_{xp}, was then computed for each trajectory as the time average of these instantaneous values:

$$\Gamma_{XP} = \frac{1}{t} \int_0^t \Gamma_{XP}(t')dt' \qquad (27)$$

The radial distribution functions *gx*(*r*) and *gw*(*r*) are defined as:

$$g_i(r) \equiv \rho_i(r) / \rho_i(\infty) \qquad (28)$$

where *i* represents water (*W*) or an additive (*X*) species. These functions provide another route to compute $\Gamma_{xp:}$

$$\begin{aligned} \Gamma_{XP} &= \langle n_X^{II} \rangle - \left\langle \left(\frac{n_X^{I}}{n_X^{I}} \right) n_W^{II} \right\rangle \\ &= \rho_X(\infty) \int g_X dV - \left(\frac{\rho_X(\infty)}{\rho_W(\infty)} \right) \rho_W(\infty) \int g_W dV \qquad (29), (30), (31) \\ &= \rho_X(\infty) \int (g_X - g_W) dV \end{aligned}$$

where each integral is over the local domain or the entire system (since *gx* - *gw* = 0 in the bulk domain).

The boundary between domains I and II must be placed far enough from the protein to ensure that it is in the bulk, yet at the smallest such distance so that statistical fluctuations in the number of molecules in the domains can be minimized. One can use the values of *gx*(*r*) and *gw*(*r*) to determine the optimal boundary. Defining Γ_{xp} as the apparent preferential binding coefficient resulting from defining the local domain as those molecules whose centers of mass lie inside a distance *r*∗ from the protein:

$$\Gamma^{*}_{XP}(r^{*}) = \rho^{\infty}_{X} \int_{0}^{r^{*}} (g_X - g_W)\frac{dV}{dr}dr \tag{32}$$

The error in Γ_{xp}, E_Γ, introduced by selecting a particular value of *r*∗ is then

$$\begin{aligned} E_\Gamma &= \Gamma^{*}_{XP} - \Gamma_{XP} \\ &= -\rho_X(\infty)\int_{r^{*}}^{\infty}(g_X - g_W)\frac{dV}{dr}dr \end{aligned} \qquad (33), (34)$$

When *r*∗ is selected properly, the surface defined by *r* = *r*∗ is entirely in the bulk solution, *gx*(*r*∗) = *gw*(*r*∗) = 1, and $E_\Gamma = 0$. Thus, selecting *r** as the minimum distance for which all *r* ≥ *r** satisfy *gx*(*r*) = *gw*(*r*) = 1 (within the error of the simulation) is optimal.

Example 3

Calculation of Constituent Group Preferential Binding Coefficients - For each simulation, up to 21 constituent group preferential binding coefficients were calculated. The 21 groups were each type of amino acid side chain present in the protein (up to 20) and the protein backbone. The "protein backbone" was defined as the -NH-CH-COO- unit, as well as the two extra protons at the N-terminus and extra oxygen atom at the C-terminus of the protein. The glycine side chain was defined as the proton bound to the alpha carbon that would be replaced by a substituent to form a different L-amino acid.

For the simulation of RNase T1 in glycerol solution, the constituent group preferential binding coefficients for the 15 individual serine residues in the protein were also calculated. For this calculation, solvent and additive molecules that were nearest to an atom in the protein that was not part of a serine side chain were not considered.

Water and additive molecules were associated with a specific constituent group by computing the distance from the center of mass of each solvent molecule to the van der Waals surface of every atom in the protein, selecting the protein atom that was nearest to the solvent molecule, and then determining to what constituent group this nearest protein atom belonged.

Example 4

Estimation of Statistical Error - The statistical error arising from computing averaged properties from a finite trajectory was estimated in the following fashion:

1. The dynamic trajectory of interest was divided into n pieces.
2. The mean of the property of interest was computed in each piece. These means were designated z_i where $i = 1 \ldots n$.
3. The standard deviation of the z_i values was computed.
4. This standard deviation was divided by n and the quotient was designated σ_m, an estimate of the error in the mean determined by time averaging the full trajectory.

The number of pieces n into which the trajectory is divided must be small enough to ensure that the means of each piece (the z_i) are statistically independent. An autocorrelation analysis (not shown) of several trajectories of $\Gamma_{xp}(t)$ data and the underlying molecular counts (n_i and n_i) indicates that a window of about 0.2 ns is sufficiently large for this to be true. Therefore, for a 2 ns dynamics trajectory, a value of $n = 2/0.2 = 10$ was used.

For long trajectories, the statistical error σ_m is roughly proportional to the inverse square root of the trajectory length. This property can be used to estimate the trajectory length required to achieve a given level of statistical accuracy after a small trajectory has been generated and analyzed.

Incorporation by Reference

All of the patents and publications cited herein are hereby incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

We claim:

1. A compound, comprising a non-protein-binding moiety (NPBM) and at least one protein-binding group (PBG).

2. The compound of claim 1, wherein the NPBM is a polyol, sugar, amino acid, or dendrimer moiety.

3. The compound of claim 1, wherein the PBG is a urea, guanidinium ion, detergent, amino acid, denaturant, surfactant, polysorbate, polaxamer, citrate, chaotrope, or acetate group.

4. The compound of claim 2, wherein the polyol moiety is a sorbitol or mannitol moiety.

5. The compound of claim 2, wherein the sugar moiety is a glucose, sucrose, or trehalose moiety.

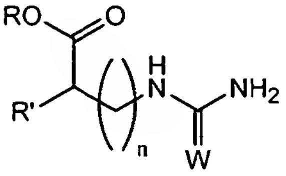
6. The compound of claim 2, wherein the amino acid moiety is an arginine betaine, proline, or ectoine moiety.

7. The compound of claim 2, wherein the dendrimer moiety is based on benzene, pentaerythritol, $P(CH_2OH)_3$, or TRIS.

8. The compound of claim 3, wherein the PBG is a guanidinium ion.

9. The compound of claim 3, wherein the PBG is sodium dodecyl sulfate.

10. The compound of claim 1, wherein the compound has formula **I**:



I

wherein:

R is an electron pair, H, alkyl, aryl, heteroaryl, aralkyl, heteroaralkyl, or an alkali metal;

R' is H, alkyl, aryl, heteroaryl, aralkyl, heteroaralkyl, or R"H_2N;

R" is an electron pair, H, alkyl, aryl, heteroaryl, aralkyl, or heteroaralkyl;

W is O, NH_2^+(halogen)$^-$, or S; and n is 1, 2, or 4-100.

11. The compound of claim 10, wherein R is an electron pair.

12. The compound of claim 10, wherein R' is H.

13. The compound of claim 10, wherein R' is R"H_2N.

14. The compound of claim 10, wherein R' is H_3N^+.

15. The compound of claim 10, wherein W is $NH_2^+Cl^-$.

16. The compound of claim 10, wherein n is 1.

17. The compound of claim 10, wherein n is 2.

18. The compound of claim 10, wherein n is 4.

19. The compound of claim 10, wherein n is 5.

20. The compound of claim 10, wherein n is 6.

21. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 1.

22. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 2.

23. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 4.

24. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 5.

25. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is $NH_2^+Cl^-$, and n is 6.

26. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is O, and n is 1.

27. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is O, and n is 2.

28. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is O, and n is 4.

29. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is O, and n is 5.

30. The compound of claim 10, wherein R is an electron pair, R' is H_3N^+, W is O, and n is 6.

31. The compound of claim 10, wherein R is an electron pair, R' is H, W is $NH_2^+Cl^-$, and n is 1.

32. The compound of claim 10, wherein R is an electron pair, R' is H, W is $NH_2^+Cl^-$, and n is 2.

33. The compound of claim 10, wherein R is an electron pair, R' is H^+, W is $NH_2^+Cl^-$, and n is 4.

34. The compound of claim 10, wherein R is an electron pair, R' is H, W is $NH_2^+Cl^-$, and n is 5.

35. The compound of claim 10, wherein R is an electron pair, R' is H, W is $NH_2^+Cl^-$, and n is 6.

36. The compound of claim 10, wherein R is an electron pair, R' is H, W is O, and n is 1.

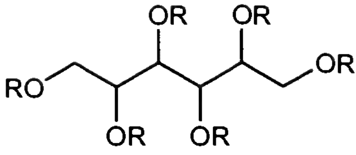
37. The compound of claim 10, wherein R is an electron pair, R' is H, W is O, and n is 2.

38. The compound of claim 10, wherein R is an electron pair, R' is H, W is O, and n is 4.

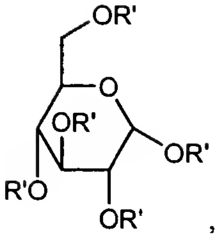
39. The compound of claim 10, wherein R is an electron pair, R' is H, W is O, and n is 5.

40. The compound of claim 10, wherein R is an electron pair, R' is H, W is O, and n is 6.

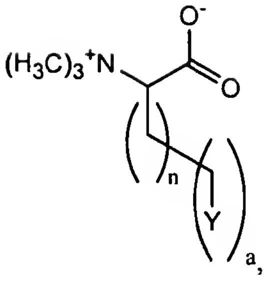
41. The compound of claim 1, wherein the compound is selected from one of the following:

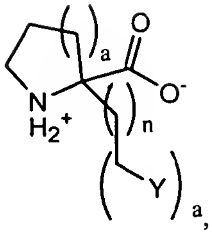


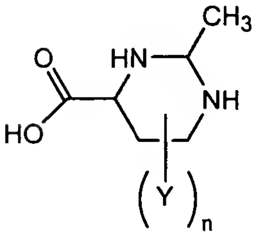
,



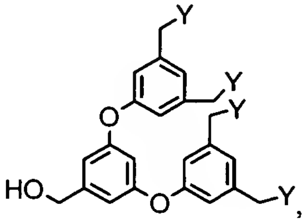
,

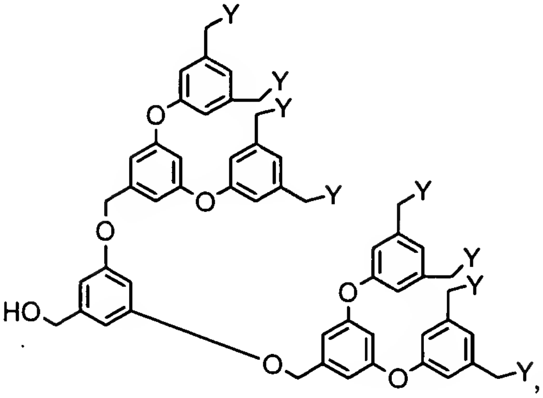




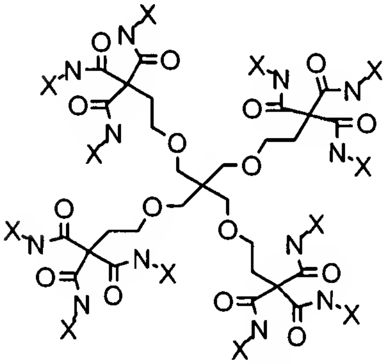


,





, or



,

wherein, independently for each occurrence,

R is H or CH_2Y;

R' is H, a sugar radical, or CH_2Y;

n is an integer from 1 to 100, inclusive;

a is 1, 2, or 3;

X is $C(CH_2Y)_3$; and

Y is a protein binding group,

wherein at least one Y is present in all compounds.

42. The compound of claim 41, wherein Y is a guanidinium ion.

43. A method of screening compounds for the property of inhibiting protein aggregation in solution, comprising:

a) computing a set of parameters utilizing molecular modeling based on compounds known to have the property of inhibiting protein aggregation;

b) applying those parameters to other compounds; and

c) choosing the compounds that meet the criteria of those parameters.

44. A method of preparing a compound having the property of protein aggregation inhibition in solution, comprising:

a) computing a set of parameters utilizing molecular modeling based on compounds known to have the property of inhibiting protein aggregation;

b) designing a compound having the property of protein aggregation inhibition in solution based on those parameters; and

c) synthesizing the compound having the property of protein aggregation inhibition in solution.

45. A method of classifying a compound as either inhibitory of protein aggregation in solution or not inhibitory of protein aggregation in solution, comprising:

a) computing a set of parameters utilizing molecular modeling based on compounds known to have the property of inhibiting protein aggregation;

b) applying those parameters to a compound; and

c) classifying the compound that meet the criteria of those parameters as inhibitory of protein aggregation in solution.

46. A method of determining the preferential binding coefficient, Γ_{XP}, of an additive in a protein solution, comprising:

a) determining the phase space trajectories of the protein, solvent, and additive using molecular dynamics;

b) calculating the distance, r, between the center of mass for both the solvent molecule and additive molecule to the protein's van der Waals surface;

c) determining the minimum distance, r*, at which no significant differences between the local (r = r*) and bulk density are observed;

d) determining which molecules lie within the distance, r*, from the protein surface and classifying these molecules as the local domain;

e) determining which molecules lie outside the distance, r*, from the protein surface and classifying these molecules as the bulk domain;

f) determining the instantaneous preferential binding coefficient, $\Gamma_{XP}(t)$, using the following formula:

$$\Gamma_{XP}(t) = n^{II}_{X} - n^{I}_{X} (n^{II}_{W} / n^{I}_{W})$$

wherein:

n^{II}_{X} = the number of additive molecules in the bulk domain;

n^{I}_{X} = the number of additive molecules in the local domain;

n^{II}_{W} = the number of solvent molecules in the bulk domain; and

n^{I}_{W} = the number of solvent molecules in the local domain; and

g) calculating the preferential binding coefficient, Γ_{XP}, as the time average of each of the values in step f) using the following formula:

$$\Gamma_{XP} = \frac{1}{t}\int_{0}^{t}\Gamma_{XP}(t')dt'.$$

47. A method of suppressing or preventing aggregation of a protein in solution, comprising the step of combining in a solution a compound of the present invention and a protein.

48. The method of claim 47, wherein the protein is a recombinant protein.

49. The method of claim 47, wherein the protein is a recombinant antibody.

50. The method of claim 47, wherein the protein is a recombinant human antibody.

51. The method of claim 47, wherein the protein is a recombinant human protein.

52. The method of claim 47, wherein the protein is recombinant human insulin, recombinant human erythropoietin or a recombinant human interferon.

53. The method of claim 47, wherein the solution is an aqueous solution.

54. The method of claim 47, wherein the protein is a recombinant protein; and the solution is an aqueous solution.

55. The method of claim 47, wherein the protein is a recombinant human protein; and the solution is an aqueous solution.

56. A method of suppressing or preventing aggregation of a protein in solution, comprising the step of combining in a solution a compound of the present invention and a protein.

57. The method of claim 56, wherein the protein is a recombinant protein.

58. The method of claim 56, wherein the protein is a recombinant antibody.

59. The method of claim 56, wherein the protein is a recombinant human antibody.

60. The method of claim 56, wherein the protein is a recombinant mammalian protein.

61. The method of claim 56, wherein the protein is a recombinant human protein.

62. The method of claim 56, wherein the protein is recombinant human insulin, recombinant human erythropoietin or a recombinant human interferon.

63. The method of claim 56, wherein the solution is an aqueous solution.

64. The method of claim 56, wherein the protein is a recombinant protein; and the solution is an aqueous solution.

65. The method of claim 56, wherein the protein is a recombinant human antibody; and the solution is an aqueous solution.

66. The method of claim 56, wherein the protein is a recombinant human protein; and the solution is an aqueous solution.

67. A method of decreasing the toxicological risk associated with administering a protein to a mammal in need thereof, comprising the steps of adding to a first solution of a protein a compound of the present invention to give a second solution; and administering to a mammal in need thereof a therapeutic amount of said second solution.

68. The method of claim 67, wherein the protein is a recombinant protein.

69. The method of claim 67, wherein the protein is a recombinant antibody.

70. The method of claim 67, wherein the protein is a recombinant human antibody. In certain embodiments, the protein is a recombinant mammalian protein.

71. The method of claim 67, wherein the protein is a recombinant human protein.

72. The method of claim 67, wherein the protein is recombinant human insulin, recombinant human erythropoietin or a recombinant human interferon.

73. The method of claim 67, wherein the first solution and the second solution are aqueous solutions.

74. The method of claim 67, wherein the protein is a recombinant protein; and the first solution and the second solution are aqueous solutions.

75. The method of claim 67, wherein the protein is a recombinant human antibody; and the first solution and the second solution are aqueous solutions.

76. The method of claim 67, wherein the protein is a recombinant human protein; and the first solution and the second solution are aqueous solutions.

77. A method of facilitating native folding of a recombinant protein in solution, comprising the step of combining in a solution a compound of the present invention and a recombinant protein.

78. The method of claim 77, wherein the recombinant protein is a recombinant antibody.

79. The method of claim 77, wherein the recombinant protein is a recombinant human antibody.

80. The method of claim 77, wherein the recombinant protein is a recombinant mammalian protein. In certain embodiments, the recombinant protein is a recombinant human protein.

81. The method of claim 77, wherein the recombinant protein is recombinant human insulin, recombinant human erythropoietin or a recombinant human interferon.

82. The method of claim 77, wherein the solution is an aqueous solution.

83. The method of claim 77, wherein the recombinant protein is a recombinant human antibody; and the solution is an aqueous solution.

84. The method of claim 77, wherein the recombinant protein is a recombinant human protein; and the solution is an aqueous solution.

Figure 1

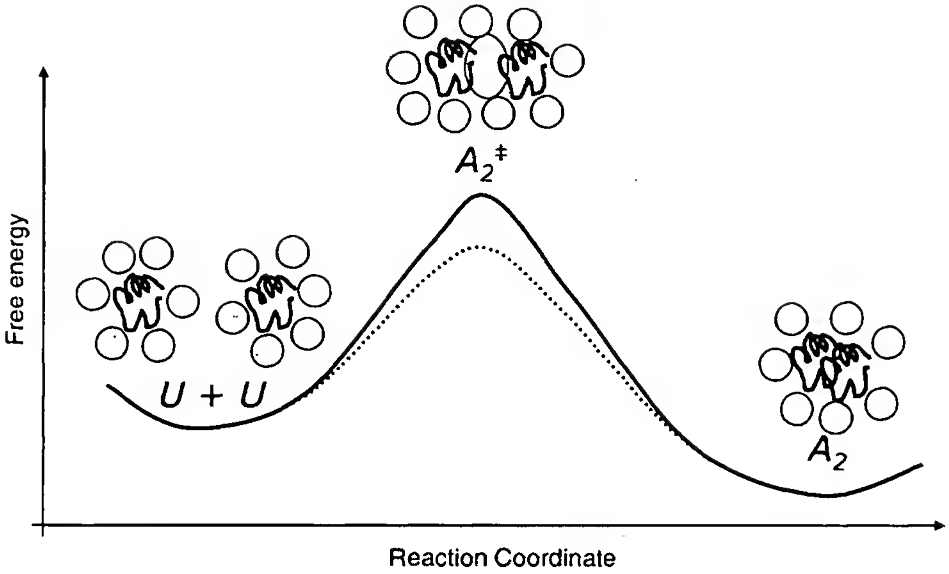


Figure 2

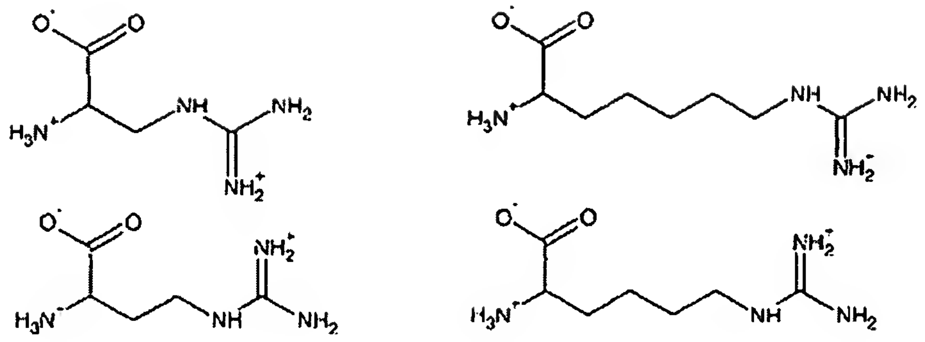

O-
O
H3N+
NH
NH2
NH2+
O-
O
H3N+
NH
NH2
NH2-
O-
O
NH2+
H3N-
NH
NH2
O-
O
NH2+
H3N+
NH
NH2

Figure 3

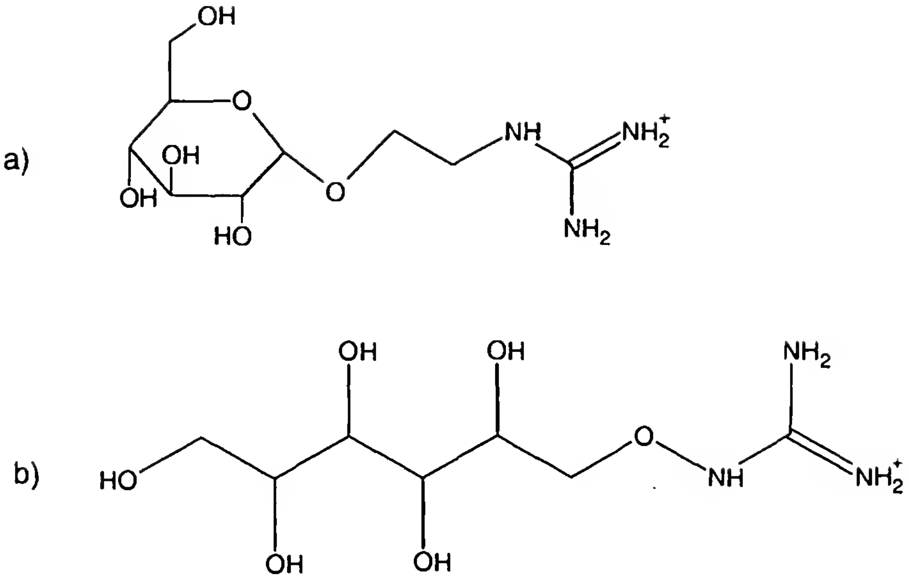


Figure 4

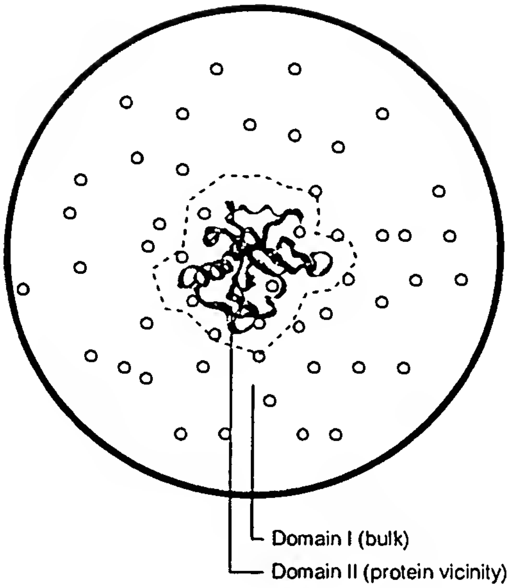


Figure 5

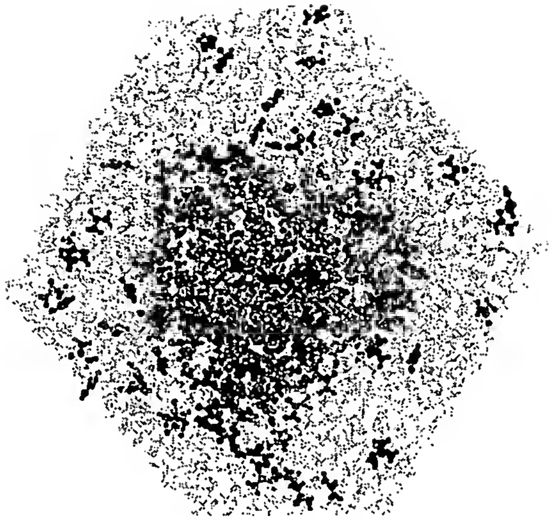


Figure 6

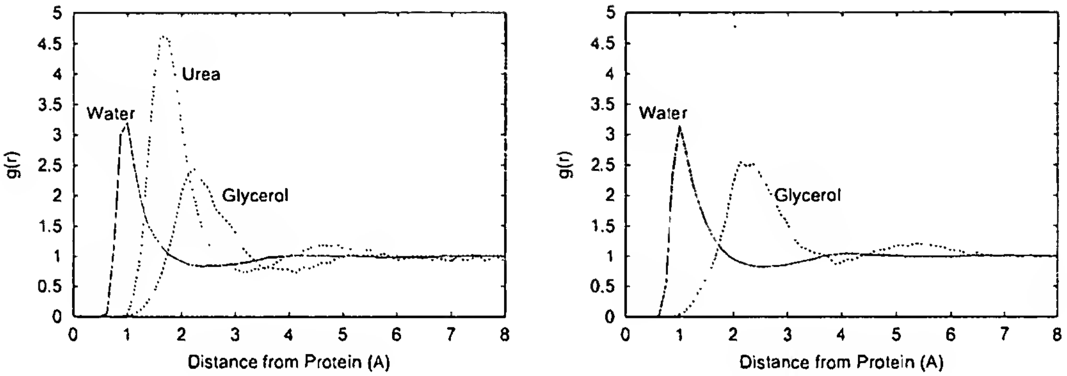

g(r)
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0 1 2 3 4 5 6 7 8
Distance from Protein (A)
Water
Urea
Glycerol
g(r)
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0 1 2 3 4 5 6 7 8
Distance from Protein (A)
Water
Glycerol

Figure 7

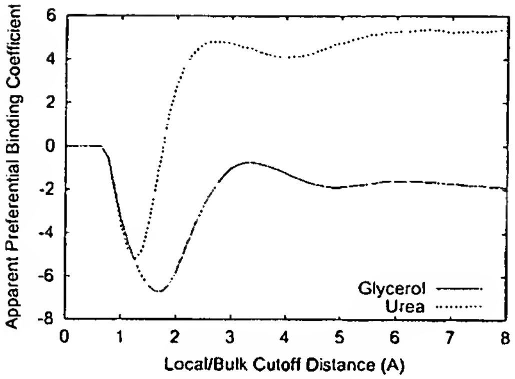

Apparent Preferential Binding Coefficient
6
4
2
0
-2
-4
-6
-8
0
1
2
3
4
5
6
7
8
Local/Bulk Cutoff Distance (A)
Glycerol
Urea

Figure 8

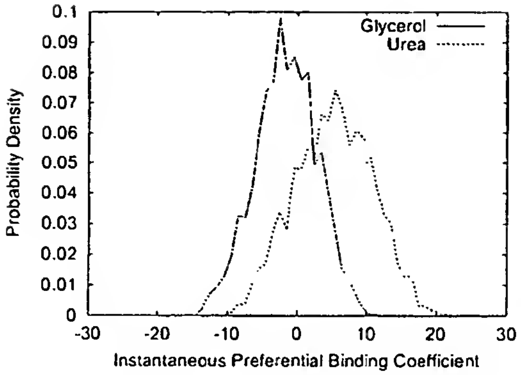

Glycerol
Urea
Probability Density
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-30
-20
-10
0
10
20
30
Instantaneous Preferential Binding Coefficient

Figure 9

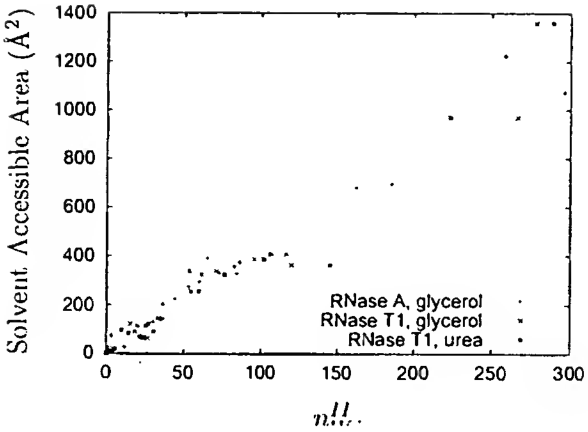

Solvent Accessible Area (Å^2)
1400
1200
1000
800
600
400
200
0
0
50
100
150
200
250
300
n^{H}
RNase A, glycerol
RNase T1, glycerol
RNase T1, urea

Figure 10

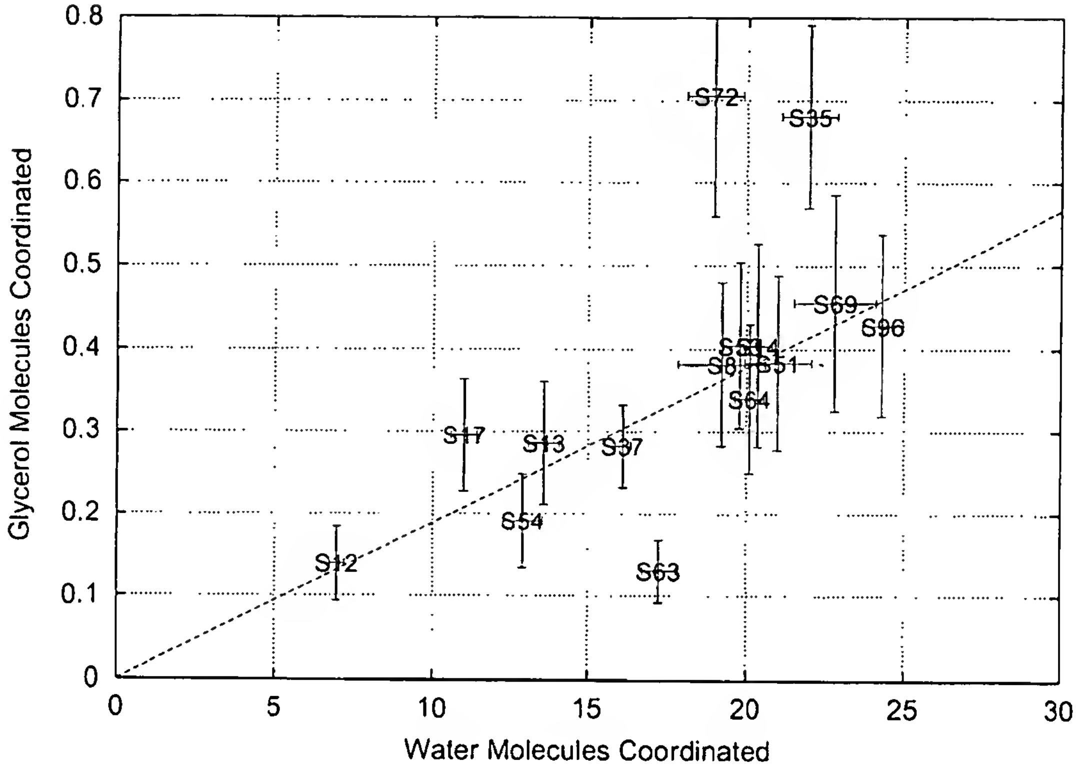

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
Glycerol Molecules Coordinated
Water Molecules Coordinated
S72
S35
S69
S96
S14
S8
S51
S64
S17
S13
S37
S54
S12
S63

Figure 11

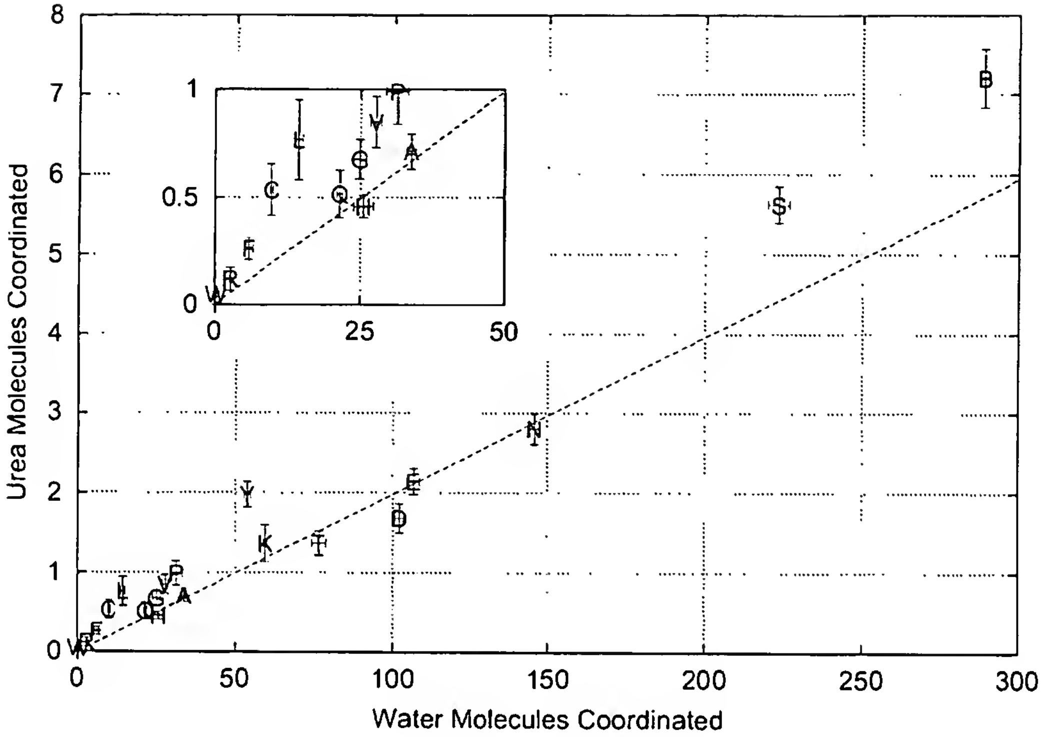

Urea Molecules Coordinated
Water Molecules Coordinated
0
1
2
3
4
5
6
7
8
0
50
100
150
200
250
300
0
0.5
1
0
25
50
W
R
F
C
L
O
G
H
V
P
A
Y
K
T
D
E
N
S
B

Figure 12

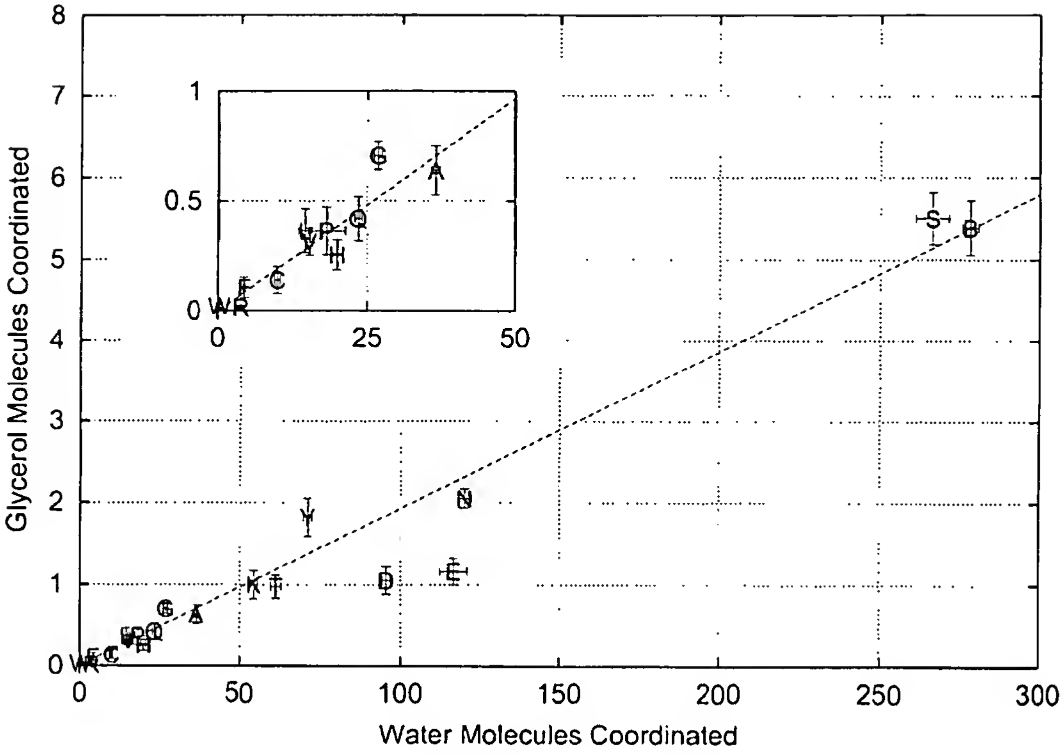
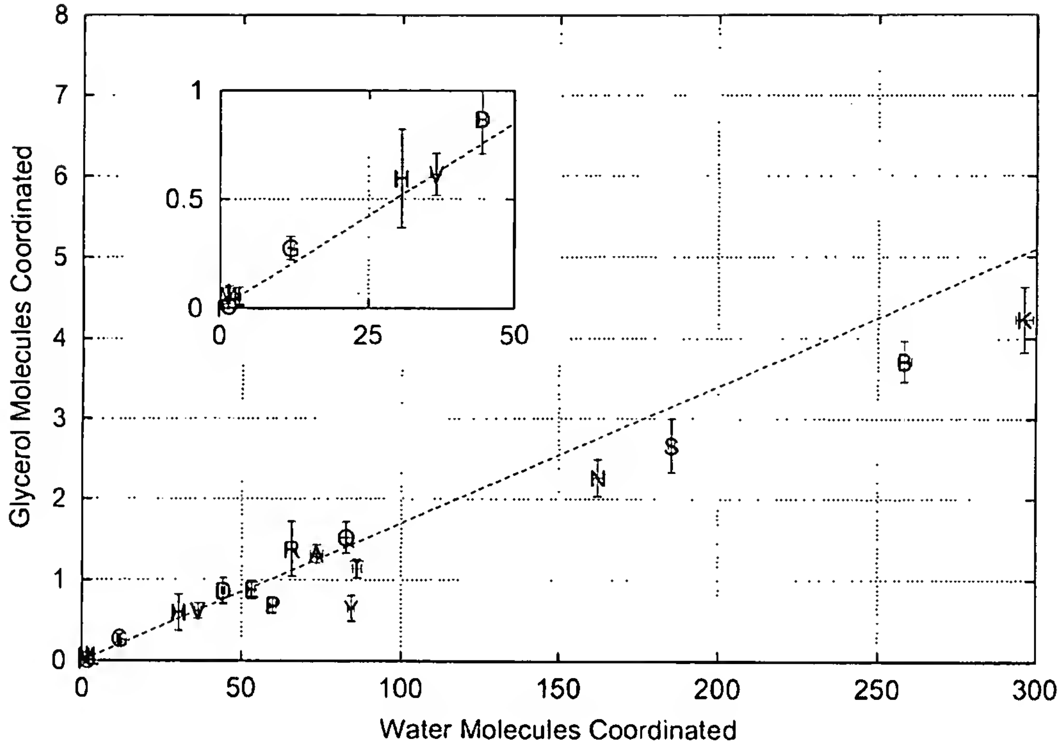

Glycerol Molecules Coordinated
Water Molecules Coordinated
0
1
2
3
4
5
6
7
8
0
50
100
150
200
250
300
W
R
F
C
V
L
P
H
Q
G
A
K
T
Y
D
E
N
S
B
0
0.5
1
0
25
50

Figure 13


Glycerol Molecules Coordinated
Water Molecules Coordinated
0
1
2
3
4
5
6
7
8
0
50
100
150
200
250
300
G
H
V
D
E
P
R
A
Q
I
Y
N
S
B
K
0
0.5
1
0
25
50
G
H
V
D